

AD-A017 007

NEW RAY METHODS IN PROPAGATION

Bernard O. Koopman

Arthur D. Little, Incorporated

Prepared for:

Office of Naval Research

1 November 1975

DISTRIBUTED BY:

**NTIS**

National Technical Information Service  
U. S. DEPARTMENT OF COMMERCE

321095

ADA017007

# NEW RAY METHODS IN PROPAGATION

prepared for  
the

OFFICE OF NAVAL RESEARCH

Contract No. N00014-72-C-0173

by

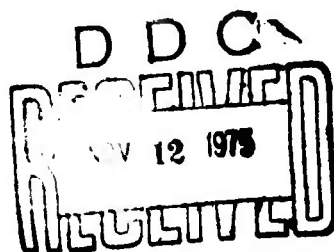
DR. BERNARD O. KOOPMAN

ARTHUR D. LITTLE, INC.

15 ACORN PARK  
CAMBRIDGE, MASSACHUSETTS 02140

Reproduced by  
NATIONAL TECHNICAL  
INFORMATION SERVICE  
US Department of Commerce  
Springfield, VA. 22151

NOVEMBER 1, 1975



DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  NEW RAY METHODS IN PROPAGATION		5. TYPE OF REPORT & PERIOD COVERED
7. AUTHOR(s)  Bernard O. Koopman		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Arthur D. Little, Inc. Acorn Park Cambridge, Mass. 02140		8. CONTRACT OR GRANT NUMBER(s)  N00014-72-C-0173
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Dept. of the Navy Arlington, Va. 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS  TASK 3
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE NOVEMBER 1975
		13. NUMBER OF PAGES 361
		15. SECURITY CLASS. (of this report)  UNCLASSIFIED
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Distribution of this document is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES  A companion report "Computation of Long Range Propagation Losses in a Duct" by R.G. Potts and C. Raisbeck describes computational methods based partly on the mathematical theories which are the subject of the present report.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Fermat Family - Travelling Waves - Hamiltonian - Phase Space - Integral Invariants - Angle Variables - Transmission Loss - Wave Equation - Canonical Transformations - Random Radiation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  The object of this report is to supply the theoretical bases for certain useful and computable approximations to the problem of transmission loss in hydroacoustic propagation. These approximations involve a knowledge of two acoustic profiles, one at the emitter and the other at the receiver and only require general assumptions of the intermediate behavior of sound speed. In this sense they form a "second approximation"--the "first approximation" being the familiar one based on the assumption of an		

DD FORM 1473

JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. acoustic profile independent of range. The ranges are long enough to involve multiple-path transmission, caustics, and the complications produced by many convergence zones. The random factors and incomplete knowledge of the medium of transmission are not only explicitly acknowledged but are made to play a central role in the discussion: they give rise to a somewhat novel model and open up possibilities of simplification of averages, etc.

The mathematical methods are based on the following three processes: First, the Hamiltonian treatment of the differential equations of the rays, together with the transcription or "lifting" into phase space of the energy density and flux quantities. This requires a detailed energy-theoretical comparison of the waves and the rays at high frequencies. Its mathematical manipulation is facilitated by the use of Hamiltonian integral invariants having statistical power-flow interpretations. Second, the physically measurable quantities are expressed in terms of these quantities. Third, a canonical transformation of phase space, based on a simultaneous solution of two Hamilton-Jacobi equations, is used to bring about the superability of the equations in the new variables, thus allowing the introduction of angle and action variables, and the expansion of the power flux in a Fourier series. Dropping all but the constant term in this series (which term depends only on the two profiles) provides the desired approximation.

In the development, many incompletenesses (and some errors) in existing treatments are noted, and attempts made to remedy them, as well as to indicate work that should still be done.

UNCLASSIFIED

1 a / SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)



NEW RAY METHODS IN  
PROPAGATION

prepared for  
the

OFFICE OF NAVAL RESEARCH

Contract No. N00014-72-C-0173

by

DR. BERNARD O. KOOPMAN

ARTHUR D. LITTLE, INC.

15 ACORN PARK  
CAMBRIDGE, MASSACHUSETTS 02140

NOVEMBER 1, 1975

## TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION	1
1. THE METHOD OF RAY TRACING	10
2. DIFFICULTIES AT LONG RANGES	15
3. WAVES AND ENERGY FLOW	19
4. TRAVELLING WAVES AND MEAN POWER FLOW	23
5. BOUNDARY CONDITIONS AND POWER WAVE-FRONTS	28
6. ASYMPTOTIC AND FERMAT FAMILIES OF TRAVELLING WAVES	35
7. A MECHANICAL ANALOGUE MODEL OF ACOUSTIC RAYS	50
8. RAYS, PARTICLE DYNAMICS AND HAMILTONIAN PHASE SPACE	56
9. DIFFERENTIAL SYSTEMS AND THEIR INTEGRAL INVARIANTS	66
10. ENERGY FLOW IN PHASE SPACE	77
11. THE STATISTICS OF ACOUSTIC TRANSMISSION	83
12. A METHOD OF STATISTICAL PERTURBATIONS	95
13. RAYS IN A VERTICAL PLANE WITH AZIMUTHAL SYMMETRY	107
14. GRAPHICAL REPRESENTATIONS AND THE SURFACE OF SECTION	117
15. CONTINUATION: THE STATISTICALLY LAMINAR CASE	144
16. TRANSMISSION LOSS WITH SECULAR NON-LAMINARITY	163
REFERENCES	231
APPENDICES:	
A. THE MATHEMATICS OF RAYS AND WAVE FRONTS: CERTAIN COMMON ERRORS	A-1
B. THE CASE OF CONSTANT ACOUSTIC PROFILE	B-1
C. INDIVIDUAL TRAVELLING WAVES	C-1

TABLE OF CONTENTS (CONTD.)

	<u>Page</u>
D. WAVES, RAYS AND POWER FLOW LINES	D-1
E. INTEGRAL INVARIANTS	E-1
F. AN INTEGRAL SOLUTION OF LINEAR DIFFERENTIAL SYSTEMS	F-1
G. TECHNICAL MATHEMATICAL SUPPLEMENT	G-1
H. PROPAGATION OVER UNDERWATER OBSTRUCTIONS AND THE "COUPLING FACTOR" FALLACY	H-1

## NEW RAY METHODS IN PROPAGATION

The ultimate objective of this study is a method of computing hydro-acoustic transmission loss at long and intermediate ranges--a method simple enough to be implemented by feasible computer programs, general enough to include the actual variations of sound speed found in nature (geographical, seasonal, and fluctuating randomly, as when influenced by turbulence, etc.); and, finally, theoretically sound in that it is derived logically from accepted mathematical physics. The immediate objective of the study is the establishment of such a basic rationale for the computation of practical answers: It is methodological and theoretical. To gain a proper perspective, we outline the generalities of the present situation, in as elementary terms as possible.<sup>1\*</sup>

The conventional methods of dealing with the propagation of high frequency radiant energy (in acoustics, optics, or electromagnetism) by means of the Huygens' construction of wave fronts and rays, meets with difficulties when the medium is heterogeneous and the ranges are long. This is precisely the case of importance in the detection of hydroacoustic signals. Caustics, path-splitting ambiguities, and other obscurities put many of the troubles into evidence.

The attempt to avoid the difficulties by the more fundamental approach through the wave equation (normal modes, etc.) is often made. While conceptually this is the correct method for explaining the obscurities in the use of rays, to be computationally effective it must go further. The use of special functions and series for solving the wave

---

\* References in the text are to the publications given in detail under the superscript numeral (assigned according to general topic).

equation requires drastically simplified assumptions regarding the sound speed  $c$  as a function of position throughout the volume of ocean in which the acoustic propagation takes place, as well as assumptions regarding the shape and nature of the ocean bottom and surface.

The difficulty--whether rays or normal modes are used--is a double one: Not only are the factors just cited highly complicated, as the limited existing observations have shown, but they are largely unknown. Even if the whole ocean bottom throughout the world were ever to be charted in minute detail, there would remain the prediction of sound speeds--a task coordinate in difficulty with predicting the weather.

Situations in which there is an irremovable residuum of unknown complexity in the determining factors affecting quantitative behavior are familiar in mathematical physics, e.g., in the kinetic theory of matter, Brownian motion, and turbulence. The successful treatments of such cases are by statistical methods: the application of the analytic theory of probability. In modern acoustics, there is a rapidly increasing view (on both sides of the "Iron Curtain") that the appropriate treatment of the acoustic problems we are discussing must be a statistical one. A profusion of books and papers written with this point of view are appearing.<sup>2</sup> Cf. e.g., Tolstoy and Clay<sup>1</sup>, Chapter 6.

The concrete question that this situation poses is whether one is forced to confine oneself to the wave equation and develop a tractable formulation of the statistics of its solutions, or whether there is still the possibility of developing a statistical treatment of the rays themselves, somehow overcoming the difficulties mentioned in

the second paragraph. Here it must be emphasized that, when applicable, a ray treatment depends on more elementary concepts and is usually computationally simpler than the direct use of the wave equation.

An essential purpose of the present report is to show how a modification of the ray treatment not only side-steps the difficulties cited in our second paragraph but provides a natural "port of entry" for statistical methods. It is based on the application of Hamiltonian theory and its recent developments (integral invariants, ergodic concepts, etc.). This application reflects a very old notion: The particle-wave duality, which was evidenced in the debate between Newton and Huygens on the "corpuscular" versus "wave theory" of light; it was exploited by Hamilton early in the last century, and returned to prominence in the de Broglie-Schroedinger wave mechanics of quantum theory.

To carry out our program along these lines and arrive at a scheme of computation of transmission loss--both practicable and realistic in view of the physical complexities just mentioned--four steps have to be taken: undertaking a mathematically sound exploration of the conditions under which rays guide the propagation of hydro-acoustic energy; a "lifting" into the Hamiltonian phase space of these energy flow-versus ray relations from the lower dimensional physical space; showing precisely how to include many of the random features of propagation in this process; and, finally, applying both old and new methods of Hamiltonian theory (integral invariants, "ergodic" concepts, and special types of canonical transformations) to establish the desired result: what we may call the "second approximation" to long range transmission loss.

If one terms the "first approximation" to transmission loss any calculation based on the assumption that the acoustic profile is independent of range, so that only the one at the emitter is used, it is natural to call the "second approximation" one making use of two profiles, the one at the receiver and at the emitter. A number of different schemes of computation of this second type have been set up, often in classified reports and based on various conceptions of the physical processes involved--and too frequently leading to quite different results. The scheme of computation to which the present study leads has been announced in different stages of its development at the Acoustical Society of America [6 Nov. 1969 (6H1): "Some Mathematical and Computational Contributions to Underwater Sound Propagation", by B.O. Koopman and G. Raisbeck; 22 Apr. 1970: (N11): "Computation of Long-Range Propagation Loss in a Duct" by G. Raisbeck; and (N12): "Propagation Over Underwater Obstructions" by B.O. Koopman]. These were not published but were presented to the U.S. Navy in the form of reports. Another essentially equivalent computational scheme, based on similar intuitive conceptions, was published by P.W. Smith, Jr., in the paper, "Averaged Sound Transmission in Range-Dependent Channels" [Journal of the Acoustical Society of America, Vol. 55, No. 6, June 1974, p. 1197]. Other similar work has been communicated to the present author in the form of a report by Tetra Tech, Inc., "A Statistical Model of Propagation in Sound Ducts", by M. Milder and L. Solomon. Both these latter and the present author presented many of these and related results in a symposium on Ray Tracing at the SACLANT Centre (La Spezia, 1971)

and are published in its unclassified proceedings.

But to our knowledge no treatment meeting the standards of classical mathematical physics is available in the literature: a mathematically sound derivation of the required algorithm from established mathematical physics under precisely stated and appropriate special assumptions. The present paper is intended as a contribution to this end--in the case of no (regular) time dependence of sound speed  $c$ . There are far more than philosophical or abstractly scientific reasons for seeking such a rigorous basis: without it, how can any navy decide which, out of the multitude of different methods submitted to it on the basis of different "physical intuitions" and "ad hoc" assumptions<sup>\*</sup>, are the ones--if any--to use in technological and tactical planning under various given conditions?

After recalling in Sections 1 and 2 the conventional method of ray tracing and its familiar difficulties, the study goes back to the wave equation and its energy density and flux vector (Section 3), normal modes and travelling waves (Sections 4,5). The treatment attempts to be elementary and keeps to the real domain while the basic issues are being faced. It is emphasized that those travelling waves which produce a net progress of acoustic power are only possible with "eigen-value degeneracy": two real solutions with the same frequency (as stated, e.g., in Morse and Ingard, 5.2, p. 206<sup>1</sup>). In Appendix C, a more precise definition on the basis of a normalization process removes such ambiguities as those recognized by Brekhovskikh (Chapter III.6, p. 229)<sup>1</sup>.

<sup>\*</sup> And sometimes even errors of fact: cf. the incorrect interpretation of Snell's law and a prevalent error in constructing shadow zones, examined in Appendix A, as well as many others noted in later Appendixes.



Section 6 considers families of travelling waves parameterized by their circular frequency  $\omega$ , studied as this increases indefinitely. The prevalent notion that at high enough frequencies the acoustic power flows essentially along rays (in the open, or near boundaries whose radii of curvature greatly exceed the wave length) requires precise interpretation. If it means that the power flow lines approach\*appropriately chosen rays and that the rate of flow of energy regarded as localized in the medium approaches  $c$  as  $\omega \rightarrow \infty$ , then it is sometimes true and sometimes not, as shown by various simple examples in Appendix D. Moreover, many conventional treatments, which show that a slow enough variation of amplitude  $A$  with increasing  $\omega$  is necessary for the validity of a ray treatment, and then assume it sufficient, are in error (counter example:  $\exp [i\omega(x \cos \omega + y \sin \omega)/c]$  which satisfies the Helmholtz equation, has unit amplitude, but approaches no limiting direction as  $\omega \rightarrow \infty$ ). This, and the essentially local or special treatments of radiation given by mathematicians, has forced us to establish necessary and sufficient conditions for the power flow-ray approach, i.e., for the validity of any ray treatment in actual media. Any family of travelling waves satisfying our conditions we call a Fermat family--for obvious reasons.

A mechanical analogue model of rays in ducts is given in Section 7, introducing the Hamiltonian formulation in Section 8; classically known properties of the differential systems to be used are outlined in Section 9, more technical material being deferred to Appendices E, F and G.

---

\*With an appropriate definition of "approach": "weak convergence" in the general cases is the necessary concept; cf. Section 6, end.

Section 10 transcribes ("lifts") the wave power flow relationships in an individual Fermat family into the Hamiltonian phase space, while Section 11, which deals with statistical ensembles of Fermat families, leads to a formulation of power flow relations by means of a continuous power flux density in that space. The elements of integral invariants (intuitively interpretable) then show that this density is constant along each ray (in the limit). In terms of this density, acoustic intensity, transmission loss, and similar desired measurable quantities are expressed. This is the theoretical key point of the present study.

The remaining five sections develop special methods of approximation to these measurable quantities under various stated conditions.

Section 12 applies conventional first order perturbation methods, regarding the perturbing term in the Hamiltonian as a statistical fluctuation.

The remaining sections confine the study to the case of azimuthal symmetry:  $c$  depends on depth and range from a symmetrical emitter only, and the propagation is in vertical coaxial planes. Section 13 particularizes the general results established above to this case. Section 14 uses a geometrical representation with the aid of the surface of section, introduced into dynamical theory at the turn of the century by H. Poincaré<sup>5</sup>, G. D. Birkhoff<sup>6</sup>, and their successors. Section 14 then examines in detail the "laminar" (or layered medium) case,  $c$  depending on depth only. Improved derivations

of the "first approximation" to transmission loss are developed, mainly as a stepping off point for the "second approximation", treated in Sections 15 and 16.

Section 15 applies the random perturbation method of Section 12 to the case of Section 14 but when the laminarity is only true on the average; it also applies methods of information theory to obtain a simplified graphical version of the "first approximation" at long ranges.

Section 16 established the "second approximation" in a deterministic but slow departure from laminarity. Briefly stated, it first makes a canonical change of variables based on solving, by line integrals, a pair of simultaneous Hamilton-Jacobi partial differential equations in the large. This transforms the problem to one in which the variables are separable, so that angle and action variables can be introduced. Then the power flux density, being periodic in the new range-like variable, can be expanded in a Fourier series. Physical reasons are given for dropping all but the constant term: this is easily calculated in terms of the two profiles (at the emitter and at the receiver) and so gives the desired second approximation. Points concerning the generalizability of the method are noted. The excessive length of Section 16 is due to its treatment of the same subject at three successively increasing levels of mathematical technicality, from largely intuitive to essentially rigorous, because of the diversity of backgrounds of the probable readers: logically it could be less than one-third of its present length.

Every attempt has been made to avoid unnecessary mathematical technicalities--to interpret as intuitively as possible those that the basic reasoning forces us to include in the text, and to relegate the rest to appendices. This has caused what may seem an excessive length of the study: if the state of the available literature had permitted, our presentation could have been greatly shortened. Thus, the only text in the English language treating Hamiltonian theory in sufficient depth for our purposes is Whittaker<sup>5</sup>; but his treatment of integral invariants does not go far enough for our needs and we have been obliged to refer to French texts (Poincaré and Cartan) and to add Appendix F on the subject, so that our presentation would be more nearly self-contained.

Appendix G deals with the mathematical problem of the existence of Fermat families as produced by radiators--in the large, where classical methods fail because of the multiple-valuedness of the eikonal and the tangled nature of the characteristic conoid. Hadamard's "elementary solution" is carried into the complex domain and made to yield the harmonic waves by contour integration. While giving the needed results in many special cases, a completely general theory requires further research.

Appendix H is essentially a reproduction of the author's paper "Propagation Over Underwater Obstructions", cited above, and in which the methods of the text are applied to such obstructions as sea mounts.

The great practical importance and the mathematical difficulty of the subject have led many workers to "measures of desperation", often leading to gross errors in mathematical reasoning. In addition to those noted above, Appendix D and H point out the one in Brekhovskikh's treatment of layered media and the "source to duct coupling" fallacy.

## 1. The Method of Ray Tracing

One standard elementary method of calculating transmission loss in underwater sound propagation takes over the methods of geometrical optics in a medium of variable but known refractive index  $n = 1/c$ , where  $c$  is the speed of propagation\*. The justification of this method is that when the wavelength is much smaller than the dimensions of the medium and the radius of curvature of the rays, the flow of acoustic energy is guided by the rays.\*\* Therefore by studying the extent that they spread out from the source of sound, the change in acoustic intensity with distance can be estimated. To make the method clear--and to show how it can lead to difficulties--we shall first examine a number of illustrations of its use.

The simplest case is that of a point emitter  $O$  in a homogeneous medium ( $n = \text{constant}$ ). All the rays are straight lines, and since the wave fronts are spheres centered at  $O$ , the rays guiding the flow of acoustic energy are those perpendicular to the wave fronts, viz., the straight lines through  $O$  (Figure 1.I). Let  $W_0$  be one of the spherical wave fronts, of radius  $r_0$ ; and let  $W$  be another, of greater radius  $r$ . On the surface  $W_0$  let a region  $S_0$  bounded by the curve  $C_0$  be traced. All the acoustic energy from  $O$  that crosses  $S_0$  will remain within the cone of rays through  $O$  and the curve  $C_0$ . It will therefore pass through the region  $S$  of  $W$ , bounded by the curve  $C$  of intersection of the cone with  $W$ .

---

\*The constant factor  $\bar{c}$  in the full expression  $n = \bar{c}/c$  is taken as unity to simplify the preliminary discussion. However, in Section 7, we set  $\bar{c} = C$ , appropriately chosen and  $\neq 1$ .

\*\*Strictly provable under certain additional assumptions given in later sections.

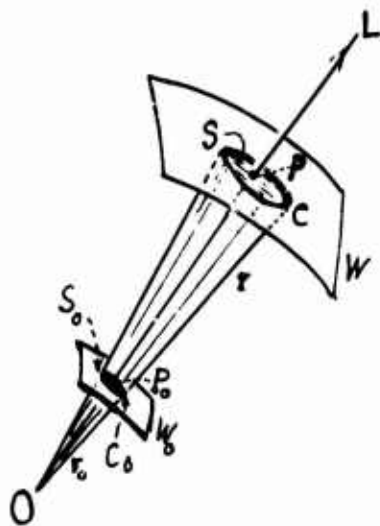


FIGURE 1.I

Spherical Spreading  
 $n = 1/c = \text{constant}$

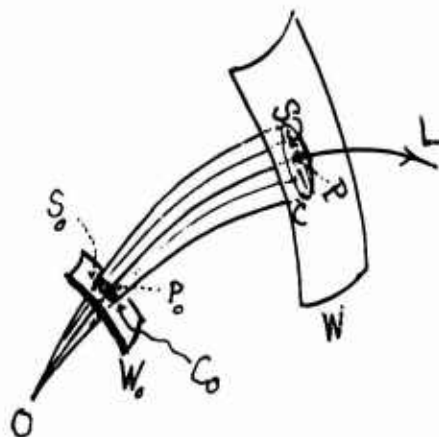


FIGURE 1.II

Deformed Spherical Spreading  
 $n = 1/c \text{ varies with position}$

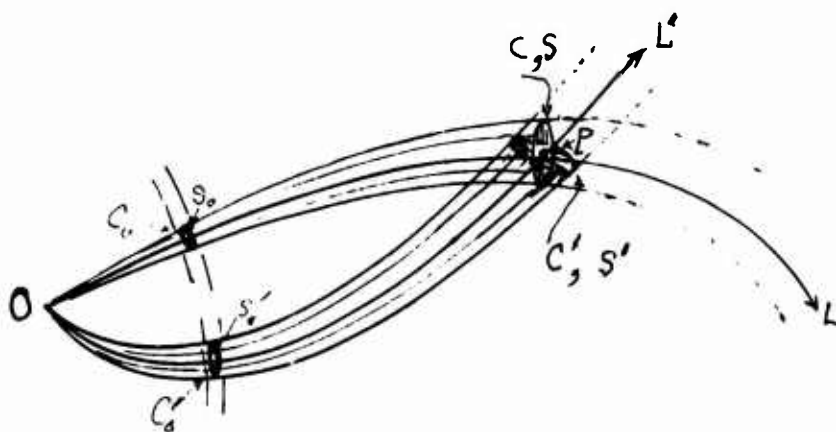


FIGURE 1.III

Multiple - Path Transmission at increased range  
 $n = 1/c \text{ varies slowly with position}$

The acoustic intensity at a point P where a ray L cuts S is the amount of energy (per unit time) intercepted by a body at this point, "subtending a unit area of the wave front." More exactly, it is the ratio of the total energy E passing through S (per unit time), divided by the area of S, the latter being regarded as arbitrarily small; more precisely, it is the limit of the ratio E/S as region S shrinks up on W to the point P.

Since we are assuming conditions under which the acoustic energy is guided by the rays, the same amount of energy E passed through  $S_0$  on its way to S, so that the acoustic intensity at  $P_0$  is the limit of the ratio  $E/S_0$ . But by the geometry of spheres,  $S/S_0 = r^2/r_0^2$ , so that the ratios of acoustic intensities at the two points obey the equation

$$(1.1) \quad \frac{\text{acoustic intensity at } P}{\text{acoustic intensity at } P_0} = \frac{S_0}{S} = \frac{r_0^2}{r^2},$$

which is the familiar inverse square law of spherical spreading. We are assuming a quasi-steady state: that any change in power output from the source O in the course of time can be neglected in this equating of the energy through  $S_0$  and through S. This is usually allowable when range r is moderate; for longer ranges, account may have to be taken of the delay time of propagation from  $S_0$  to S.

Now consider the more usual case in hydroacoustic propagation, that the index  $n = 1/c$  varies enough over the region of propagation so that the rays are no longer straight lines but gently curved. (We are postponing the case of reflection.) Figure I-II shows the modification needed in the shapes shown in Figure 1. With emission from the point O, there will still be a set of wave fronts, such as  $W_0$  and W, perpendicular to the rays through O (by the Theorem of Malus). The first equation

in (1.1) will apply; but since the areas  $S_0$  and  $S$  are not spherical, the second equation will not be true: The inverse square law of propagation no longer holds. Consequently, before the effect of spreading on transmission loss can be calculated by equation (1.1), the value of the spreading factor  $S/S_0$  ( $S_0$  small) must be found. More precisely, the limit of this ratio must be evaluated as  $S_0$  shrinks to point  $P_0$ --thereby making  $S$  shrink to  $P$ --where  $P_0$  and  $P$  lie on the same ray  $L$  through  $O$ . We denote this limit by  $\delta S/\delta S_0$ .

The method of "ray tracing" offers a graphical evaluation of the spreading factor. In its simplest terms, the method consists in constructing (e.g., with the aid of machine computation) a large number of rays emanating from  $O$ , and closely spaced in their directions from  $O$ , as evenly as convenient. If  $N(S_0)$  of these rays cut through  $S_0$ , the number  $N(S)$  cutting through  $S$  must be exactly the same, by construction of  $S$ :  $N(S) = N(S_0)$ . Therefore this common number behaves analogously to the energy: its density (per unit area normal to  $L$ ) is  $N(S_0)/S_0$  at  $P_0$  and  $N(S)/S$  at  $P$ ; hence the ratio of densities of rays at the two points  $P$  and  $P_0$  is the latter quotient divided by the former. After cancelling the common numerator, we get  $S_0/S$ , the desired term in (1.1), or reciprocal of the spreading factor. Since the ray density at any point can be read off approximately once enough rays have been traced, we are given a solution of our problem:  $S_0/S = \text{ray d. at } P / \text{r.d. at } P_0$ .

The practicality, accuracy, and reliability of the above method must all be carefully examined. We have begun with the simplest situation in which, as shown in Figure 1.II we are close enough to the emitter so that through each point  $P$  one and only ray passes, emanating



from 0. For longer ranges (beyond the onset of convergence zone phenomena) this is no longer true: multiple-path transmission sets in (Figure 1. III), together with reflections, caustics, and other complications. Their study will be undertaken in Section 2 (the mathematics, in Appendices A and B).

The accuracy and reliability of the method depend on the correctness with which the rays can be determined; whereas its practicality--indeed, its economy--is a relative matter, requiring a comparison with other methods of answering the same questions and starting from the same data. All applications of rays to the study of transmission loss--whether the rays are actually traced, computed numerically, or merely enter into the mathematical reasoning leading to numerical results--have the same starting point: an assumed knowledge of the sound speed  $c$  throughout the region of propagation (or equivalently, of the acoustic index  $n = 1/c$ ); and the differential equations determining the rays, and expressing Fermat's "principle of least time". The basic elementary facts will be recalled in their general setting in Appendix A, with specific detail given in important special cases in Appendix B.

## 2. Difficulties at Long Ranges

All the formulas and methods for calculating acoustic intensity (Section 1 and Appendix A and B) imply single path transmission: the situation of Figures 1.I, 1.II, but not of Figure 1.III. Everything has been based on the smooth wave front  $W$  which is the surface of constant phase or travel time  $t$ , orthogonal to all the rays issuing from the source  $O$ , and progressing with increase of time according to Huygens' construction.

What becomes of the construction and reasoning when longer ranges are considered, giving rise to multiple-path transmission, with such attendant complications as caustics? Such factors will surely enter at ranges beyond the first convergence zone, e.g., much beyond 35 miles. Even without the added complications of reflections or absorptions from the surface or bottom, the difficulties mentioned will occur. They have long produced practical complications and obscurities in the various computational techniques (ray tracing and analytic formulas), and many attempts have been made to deal with them by ad hoc modifications of the standard computational procedures. But before a firm basis for reliable and efficient methods can be laid, it is necessary to go back to first principles and examine the physical relation between rays, wave fronts, and the propagation of power.

Let us see geometrically how the difficulties occur, by following the progress of a wave front  $W$  by means of Huygens' construction. Figure 2-1 shows (in plane section) a front which at first has no singularities, but in which the sound speed  $c$  is greater above and below a central axis through the emitter  $O$ . In the early positions of the front,

such as  $W_1, W_2, W_3$ , it has the appearance and is what we would expect Huygens' construction to give when the "elementary (spherical) wavelets" have larger radii away from the central axis. In such later positions as  $W_4, W_5, W_6$ , the parts off the axis have more than caught up with those near the axis, and the front is converging. By the time  $W_6$  is reached, the radius of curvature on the axis is so nearly equal to that of the elementary wavelet there that the Huygens' construction will show the formation of an angular point  $F$  on the wave front  $W_7$ . The radii of curvature of the wave fronts from  $W_6$  to  $W_7$  approach zero on the axial ray. Consequently the family of rays between  $P_1F$  and  $P_2F$ , all being orthogonal to the wave fronts, will pass through the angular point  $F$  of  $W_7$ . Since the energy traversing the piece  $S_1$  of  $W_1$  will come to a focus at  $F$ , which is a piece of  $W_7$  of area zero, the energy density at  $F$  is infinite. Obviously we are dealing with a situation in which the ray approximation to the behavior of the acoustic propagation is inadequate. Thus such a simple ray tracing method as the one described in Section 1 will not tell us how to calculate the acoustic intensity beyond  $F$ .

A similar situation is illustrated schematically in Figure 2.11 which shows a pencil of rays having a caustic  $C$ . The wave front converges to the arc of tangency with  $C$ , as shown in  $A_1B_1, A_2B_2, A_3B_3$ ; it is then reversed and diverges from this arc, as shown in  $A_4B_4, A_5B_5, A_6B_6$ . Its evolution between  $A_3B_3$  and  $A_4B_4$  is not shown in the figure because of difficulties of scale; but it clearly folds over, temporarily acquiring a cusp in transit. This is because at each point  $T$  (not shown) where one of the rays touches  $C$ , one position of the wave is formed by joining two curves,  $AT$  and  $TB$ ; and at  $T$  both curves, being perpendicular

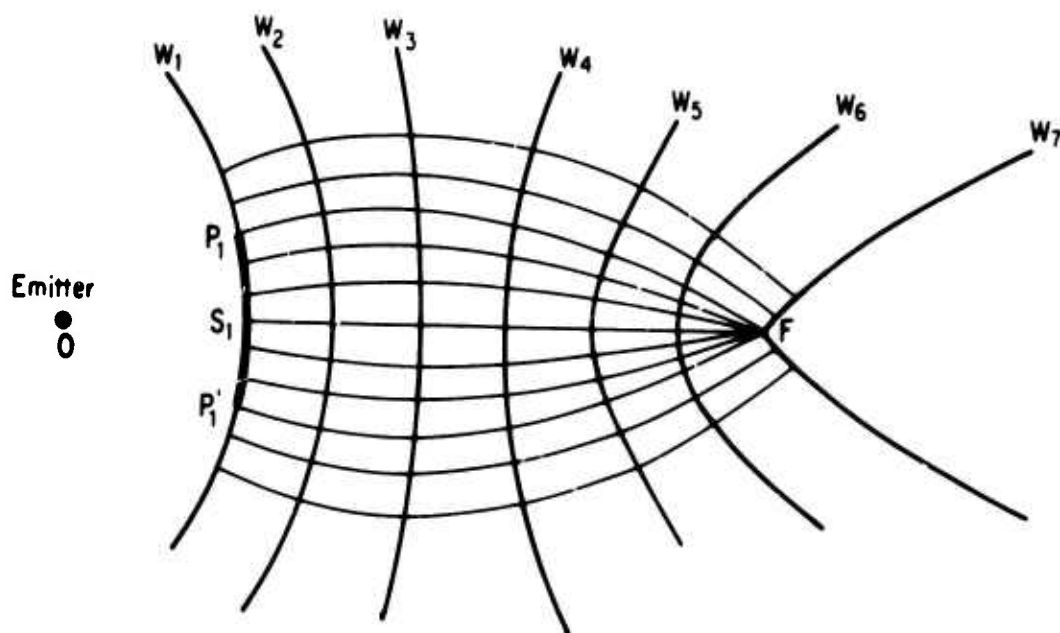


FIGURE 2-I DEVELOPMENT OF SINGULARITY F ON WAVE FRONT  $W_7$ .

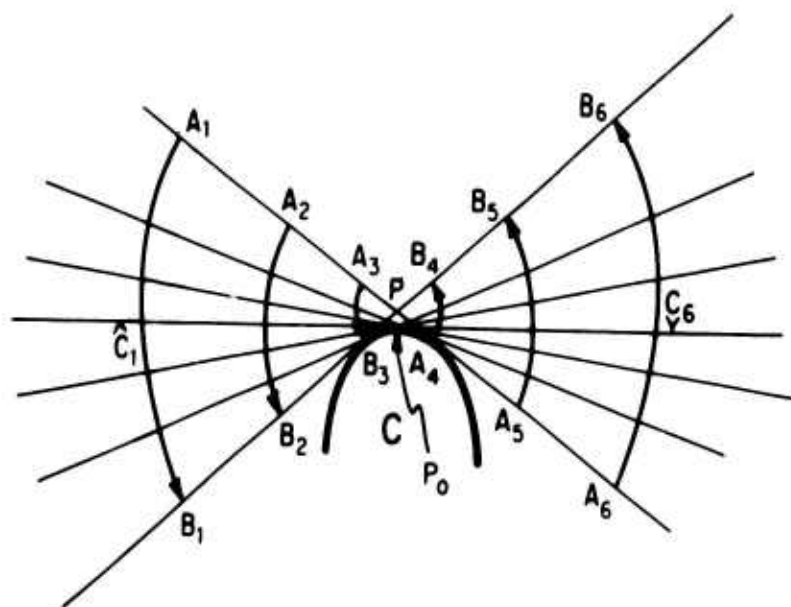


FIGURE 2-II - SINGULARITY AND REVERSAL OF WAVE FRONT  $A_i B_i$  AT CAUSTIC C.

to the ray, are also perpendicular to C. Evidently the caustic C is a locus of singularities in the wave front W; and the energy density is infinite on C. In general, caustics are loci of wave front cusps.

Analytically, the progressing wave front W is the locus of the equation  $S(x,y,z) = t$ , for the succession of increasing values to t. Here  $S(x,y,z)$  is the time taken for a signal to pass from the point of emission (O in Figures 1.I, II, III) to the point  $(x,y,z)$  of the medium-- along the "path of least time" (the ray) connecting these points. Evidently when, as in Figure 1.IV, more than one such path joins these points,  $S(x,y,z)$  is multiple-valued. Since this function was single-valued close to O (the situation in Figures 1.I and II), as  $(x,y,z)$  moves away from O, it must pass through a locus of critical points at which the single-valued S becomes multiple-valued: this "branching locus" is precisely the caustic, as Figure 2.IV shows geometrically, or the focal point, as in Figure 2.I. With such a singularity appearing in W (or S), there can be no simple form of the Huygens' construction.

### 3. Waves and Energy Flow

Section 1 has been based on the assumption that, at the signal frequencies of interest, the rays guide the flow of acoustic power. Section 2 has noted the difficulties of the conventional ray and-wave-front construction at long ranges in the case of a sound speed varying in space. Inasmuch as propagation does in fact occur under such conditions without observable difficulty, the basic trouble must be, not with Nature, but with the Huygens simplified description of it. This forces us back to first principles: the wave equation and the concomittant energy density  $E$  and its flux vector  $\vec{F}$ .

Every simple acoustic disturbance corresponds to a function  $\psi$  of time and spacial position, satisfying d'Alembert's wave equation  $c^2 \nabla^2 \psi = \psi_{tt}$  or one of its generalizations,<sup>1</sup> Using rectangular coordinates  $(x,y,z)$ , and subscripts to denote partial derivatives, the appropriate wave equation for heterogenous media is

$$(3.1) \quad \frac{1}{\rho} [(\rho \psi_x)_x + (\rho \psi_y)_y + (\rho \psi_z)_z] = \frac{1}{c^2} \psi_{tt}$$

The wave function  $\psi$  denotes the acoustic velocity potential, although (3.1) would also be valid for acoustic pressure, on replacing  $\rho$  by  $1/\rho$ . The space density of energy  $E$  and its flux vector  $\vec{F}$  are given by the vector equations ( $\nabla$ , the gradient operator).

$$(3.2) \quad E = (\rho/2) [|\nabla \psi|^2 + n^2 \psi_t^2], \quad \vec{F} = -\rho \psi_t \nabla \psi$$

Here  $\rho$  is the "quiet" density of the medium. Since the essence of the phenomena under present study is the effect of spacial variation of  $\rho$  and  $c$

(and  $n$ ), the left side of (3.1) cannot be replaced by  $\nabla^2 \psi$ --a fact which seems to have been widely disregarded.

The following theorem is basic; it is easily derived from (3.1) and (3.2) by obvious differentiations and reductions:

**Theorem:** When the quantities  $\rho$ ,  $c$  (and hence  $n$ ) are independent of the time  $t$ , the transmission of acoustic energy is conservative, obeying the "equation of continuity".

$$(3.3) \quad \frac{\partial E}{\partial t} + \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 0.$$

This means that, for a given wave  $\psi$  satisfying (3.1), we may picture the transmission of acoustic energy just as we do the transmission or flow of matter (mass) in a fluid of density  $\rho = E$  and velocity components  $u$ ,  $v$ ,  $w$ , defined by the equations

$$(3.4) \quad \rho u = Eu = X, \quad \rho v = Ev = Y, \quad \rho w = Ew = Z.$$

As in the case of real fluids, the lines of flow or trajectories are the paths in geometrical space of points  $(x, y, z)$  which are functions of time  $t$  (and initial positions)--the functions satisfying the differential equations

$$dx/dt = u = X/E, \quad dy/dt = v = Y/E, \quad dz/dt = w = Z/E.$$

At each instant  $t$ , the directions of motion will have components  $u$ ,  $v$ ,  $w$  (or, equivalently,  $X$ ,  $Y$ ,  $Z$ ). The curves which at that particular instant are tangent, at each of their points, to these

directions are the stream-lines; they satisfy (for the fixed  $t$  in question) the differential equations

$$\frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w} ;$$

or equivalently,

$$(3.5) \quad \frac{dx}{X} = \frac{dy}{Y} = \frac{dz}{Z} .$$

In general, these will depend on the particular value chosen for  $t$ : the stream-lines will vary with the time.

In the important special case of steady flow in hydrodynamics, the velocity components  $u, v, w$  are independent of the time  $t$ ; or, equivalently, the ratios  $X:Y:Z$  are independent of  $t$ . Then the stream-lines coincide with the trajectories, forming a fixed system of curves, one and only one through each point in space. Furthermore, as a direct consequence of the equation of continuity, no matter (energy, in the radiation analogy) can pass through a surface made up of these fixed stream-lines. Such a surface could be generated by moving a point  $P$  along an open curve  $C$  (not tangent to a stream line), and letting the unique stream-line through  $P$  sweep out a sheet-like surface. If, on the other hand,  $C$  were small and closed, the surface would be a tube; if  $S_0$  and  $S$  are two surfaces closing this tube (as in Figures 1.I and 1.II), the mass of fluid (or energy) entering through  $S_0$  is the same as that leaving through  $S$ .

Since we have in Section 1 been considering the rays as guiding the flow or power, it is natural to suppose that, at least for broad classes of waves  $\psi$ , the corresponding flow curves of acoustic energy are



fixed, corresponding to a steady flow, and that they exactly (or approximately) coincide with a corresponding class of acoustic rays.

Unfortunately for the simplicity of this picture, it can be shown mathematically that no real-valued  $\psi$ , satisfying the wave equation and giving rise to these stationary flow patterns, can exist—without the energy density increasing indefinitely. Only with complex-valued  $\psi$ —equivalently, with pairs of real-valued waves—can we establish such a steady flow. In the remainder of the text, we shall consider only real-valued wave functions for the sake of ease of physical conceptualism. In Appendix C, the subject is re-examined with the aid of more technical mathematics. There the concept of "travelling wave" is examined in some mathematical generality.

To pass from the wave to the ray picture through the equations (3.2), three steps have to be taken: First, we must consider waves  $\psi$  which represent stationary solutions, i.e., of the form  $\psi = \phi T$ , where  $\phi$  depends on spacial position only, and  $T$  on time only (normal modes). Second, we have to find pairs of such solutions, complying with the appropriate boundary conditions (equivalently, a single complex-valued solution). Third, after substituting a linear combination of such a pair of solutions into (3.2), the time-average of the resulting  $\vec{F}$  must be calculated. Then and only then do we obtain a steady flow, approximated by a class of rays.

#### 4. Travelling Waves and Mean Power Flow

The concept of travelling or progressing waves is familiar to electrical engineers in the study of radiation (e.g., from antennas) and of wave guides. In such cases the signal speed  $c$  is either constant or varies only at limited loci, such as boundaries between different media; cf. J.A. Stratton<sup>3</sup>. More generally, the speed  $c$  varies continuously throughout the whole region of propagation. The latter case has been treated when  $c$  is assumed to have special simplified forms, using special functions. For more general  $c$ , mathematical treatments have been given; cf. Courant & Hilbert<sup>3</sup>; but with results that are local and inapplicable to long ranges. Yet it is under precisely such conditions that many of the most important cases of signal transmissions in the ocean take place. To fill the gap in the existing treatments, this and the next two sections outline the general concepts of travelling waves in a form needed for these applications, leaving the more complicated mathematics to Appendix C

Throughout this whole study we continue to postulate the independence of sound speed  $c$  of time  $t$ ; and likewise for  $\rho$ .\*

The most important acoustic disturbances  $\psi$  in underwater detection are periodic in the time (or made up of extended pieces having this property). By a basic theorem of Fourier analysis, every such periodic function (under the usual physical assumptions of regularity) can be written as a convergent Fourier series:

$$(4.1) \quad \psi = \sum_{n=0}^{\infty} [A_n \cos(n\Omega t) + B_n \sin(n\Omega t)] ,$$

---

\*cf., however, Tatarski.<sup>2</sup>

where the Fourier coefficients  $A_n$  and  $B_n$  are independent of  $t$  but are functions of position (e.g., of the rectangular coordinates  $(x,y,z)$ ), and where  $\Omega$  is a constant, the angular frequency of the "fundamental note"; its frequency is  $\Omega/2\pi$  and its period  $2\pi/\Omega$ .  $2\pi c/\Omega$  is the approximate wave-length (exact only when  $c$  is constant). Similarly for the  $n$ 'th harmonic, of angular frequency  $\omega = n\Omega$ : its frequency  $\nu = \omega/2\pi$ , period  $= 1/\nu$ , and approximate wave-length  $c/\nu$ .

The substitution of the expression (4.1) into the wave equation, followed by formal manipulations, and reasoning familiar in Fourier analysis, show that the coefficients in (4.1) all satisfy Helmholtz' equation in its generalized form:

$$(4.2) \quad \frac{1}{\rho} \nabla \cdot (\rho \nabla \phi) + \frac{\omega^2}{c^2} \phi = 0, (\phi = A_n \text{ or } B_n, \omega = n\Omega).$$

These, multiplied by the corresponding sinusoid, are normal modes: Thus (4.1) expresses our periodic wave  $\psi$  as a sum of normal modes whose frequencies are integral multiples of a fundamental frequency.

The substitution of the expression for  $\psi$  in (4.1) into the equations (3.2) for the space density and vector flux of energy produces a sum of products of sines and cosines of various integral multiples of  $\Omega t$ , all multiplied by factors depending on spacial position only. The resulting sum, even in the simplest cases, is highly complicated, revealing the fine details of the diffraction patterns, and all the surges of energy within them.

In contrast to such complexities, if we seek what is important in the practical problems of hydroacoustic transmission, we shall naturally look for the behavior on the average. To find the time

averages of  $E$  and  $\vec{F}$  in the present case, we have but to take the term-by-term integral over one full period  $2\pi/\Omega$  of the results of the substitutions described above, and then divide the sum of integrals so obtained by this period. By elementary trigometric integration, all the resulting terms are zero except those integrals that involve  $t$  through the square of one and the same sinusoidal term; and the value of all these is  $1/2$ .

After carrying out this process, two general results are obtained:

1. The mean energy density  $E'$  is the sum of the mean energies contributed by each of the brackets in the series in (4.1); and likewise the total mean flux  $\vec{F}'$  is the vector sum of the mean flux vectors contributed by each of these brackets.

2. For each bracket, which we write simply as  $A \cos \omega t + B \sin \omega t$ , the mean energy density  $E'$  and flux vector  $\vec{F}'$  are given by the expressions

$$(4.3) \quad E' = \frac{1}{2} \left[ \frac{1}{2} (|\nabla A|^2 + \frac{\omega^2}{c^2} A^2) + \frac{1}{2} (|\nabla B|^2 + \frac{\omega^2}{c^2} B^2) \right] \rho$$

$$(4.4) \quad \vec{F}' = \frac{\omega}{2} (A \nabla B - B \nabla A) \rho.$$

Since  $t$  has been integrated out in the averaging process, all the quantities in these two formulas are independent of the time. Therefore, if we replace  $(X, Y, Z)$  in (3.5) by the components  $(X', Y', Z')$  of  $\vec{F}'$ , we obtain the stationary flow picture described in Section 3, but rejected as a general possibility for hydro-acoustic power--but now seen to be valid after the averaging. Clearly the equation of continuity (3.3) is true for the accented quantities, since  $\partial E' / \partial t = 0$  and (4.4) shows that the divergence  $\nabla \cdot \vec{F}' = 0$ .

The concept of the travelling wave applies only to those special vibrations which, in addition to the periodic property expressed in (4.1), actually give a net transfer of energy with the passage of a full period of time. From what we have seen, they are composed of pairs of "monochromatic" waves, combined together in the form  $\psi = A \cos \omega t + B \sin \omega t$ , the spacial coefficients satisfying (4.2); and with  $E'$  and  $\vec{F}'$  given by (4.3) and (4.4). Our detailed examination will be confined to such "elementary" waves. When such pairs A and B of solutions of (4.2) exist, any other pair  $A'$ ,  $B'$ , formed by taking independent linear combinations with constant coefficients of the original pair serves equally well, and when the determinant of the linear transformation is unity, it gives the identical mean flux of energy. This flexibility will be used, as explained in Appendix C, to "normalize" and "orthogonalize" our pair: this will have the physical consequence that, in a certain sense, we shall have an elementary wave producing the maximum energy flow for the least energy density. Such pairs shall constitute the travelling waves to be considered; they include all the special cases studied in conventional radiation theory.

In most of the classical problems involving normal modes of vibration, a bounded region is assumed and the boundary conditions determine the solution of a given frequency uniquely: no two linearly independent eigenfunctions can belong to the same eigenvalue  $\omega^2$ . Such cases are those of the vibrating string or elastic membrane clamped at its boundaries (one and two dimensional vibrations); more generally, cases where the (1, 2, or 3-dimensional) region is simply connected, and at every point on the boundary either  $\psi$  or its normal derivative or a

homogeneous linear combination of them is zero. Under such conditions, (4.4) shows that  $\vec{F}' = 0$ , so there is no displacement of power on the average: we do not have travelling waves.

There are exceptions in the above cases: The homogeneous square drumhead has pairs of linearly independent normal modes, and a study of the situation with the aid of (4.4) shows that there is then a net average circulation of energy about the center of the drumhead, around which the flow-lines are closed curves. When the region of vibration is multiply connected, ring-shaped for example, the above situation is the rule rather than the exception. We might call the vibration a circulating travelling wave in cases of this type. They may have applications to wave guides, but make no evident contribution to long range propagation, and shall be excluded from further study here.

In concluding this section, it may be noted that the conventional treatments usually start by assuming the existence of a complex valued normal mode. Since the real and imaginary components of such a complex function constitute a pair A, B of real fundamental modes of the same frequency (we have but to write  $C = A + iB$ ), the assumption in question is valid only in certain cases and not in others. \* Moreover, not only the physical pictures but the logical relationships of the possible assumptions are kept clearer by remaining in the domain of reals--even though many formulas are simpler in complex notation.

---

\*Some authors are clearer than others on this subject. See references to Morse & Ingard<sup>1</sup> in Appendix D for a realistic treatment.

## 5. Boundary Conditions and Power Wave-Fronts

It has been noted above that the existence or non-existence of travelling waves, and their nature when they do exist, is fundamentally dependent on the conditions set at the boundary. This seems natural, since if the boundary conditions allow no energy to cross it on the average, there can at most be circulating travelling waves; and in a simply connected region, even these are exceptional (requiring coincident eigenvalues). Such classical boundary conditions as assigning the values of the function, or of a linear combination of it and its normal derivative, can be seen by an easy application of (4.4), to give a zero mean flux of power across the boundary: The mathematics then shows that the solution of (4.2) is in general unique, etc., thus illustrating the facts concerning travelling waves. The practical question facing us here is what conditions to assume at the physical boundaries of the ocean.

The ocean regions of relevance to the present study have three types of boundary: physical boundaries of the water mass involved, i.e., its surface, bottom, or emerging land masses; emission surfaces out of which a mean flow of energy can take place, determined by the enclosed source; and reception surfaces, containing the receiver into which the energy enters. The problem is indeterminate--as it should be: If, first-ly, we imposed boundary conditions of the classical type mentioned above on all the boundaries, the real  $\phi$  would be uniquely determined and there could be no mean energy flow. If, secondly, conditions on the emission and reception boundaries are less definite, there would be infinitely many mathematical solutions of the acoustic wave problem. They might

all have "physical meaning" (i.e., correspond to possible vibrations); but only an (infinite) sub-set would be relevant to our study of signal transmission. These are the waves that we wish to single out here, by applying the criteria set forth here and in the next section.

As we have seen, the lines of mean flow of energy for an elementary travelling wave are the solution curves of the differential system

$$(5.1) \quad \frac{dx}{x'} = \frac{dy}{y'} = \frac{dz}{z'}, \quad \vec{F}' = (x', y', z')$$

where  $\vec{F}'$  is given by (4.4). Their geometrical nature will depend, in an essential way, on the loci of points in space where simultaneously  $A = 0$  and  $B = 0$ . We can say that "in general" the individual locus of  $A = 0$  on the one hand and  $B = 0$  on the other are surfaces, and that the simultaneous locus, which is their intersection, is made up of curves. Of course in equally general cases, one or both of these individual loci may not exist, or it may be made up of points or curves; furthermore, the two individual loci may not intersect, or may do so only at isolated points. But whatever the dimensionality of the locus of the simultaneous equations  $A = 0, B = 0$  it will be called the singular locus of (5.1).

A possibility not mentioned above is that the singular locus have two-dimensional parts, i.e., that the individual loci  $A = 0$  and  $B = 0$  have a surface in common, across which no mean power flows, by (4.4). Then such a surface is an "internal boundary" made up of flow curves (determined by (5.1) and its consequences). This theorem, which is a consequence of Helmholtz' equation (4.2), is proved in Appendix C,



where it is also proved that it creates no ambiguities in the flow direction  $(X':Y':Z')$ , nor in the limiting values of  $\theta$  in (5.3), which differ by  $\pi$  as a regular point on it is approached from the two sides.

Starting from any elementary travelling wave (under the restriction just mentioned), to any point  $(x,y,z)$  of the part of the ocean considered there corresponds a point  $(u,v)$  in a plane, given by the equations

$$(5.2) \quad u = A = A(x,y,z), \quad v = B = B(x,y,z).$$

To points on the singular locus corresponds the origin  $(0,0)$  in the  $uv$ -plane; to every other point, there corresponds a point  $(A,B) \neq (0,0)$ . Such a point defines a distance  $R = (A^2 + B^2)^{1/2}$  to the origin, and, apart from integral multiples of  $2\pi$  radians, an angle  $\theta$  whose sine and cosine are  $A/R$  and  $B/R$  (or their limits on the 2-dimensional singular loci):

$$(5.3) \quad A = R \cos\theta, \quad B = R \sin\theta.$$

Clearly these polar functions of  $(x,y,z)$  have all the regularity properties of  $A$  and  $B$  (continuous differentiability, etc.), provided  $(x,y,z)$  is not on the singular locus. On the other hand, while  $R$  is single valued,  $\theta$  may well be infinitely multiple valued. This multiplicity may be generated when exclusion of the singular locus from the region of propagation renders the latter multiply connected. Then the point  $(x,y,z)$  can move along a closed curve threading the excluded singular locus: if, as may happen, the corresponding point  $(A,B)$  in the  $uv$ -plane goes around the origin  $(0,0)$  e.g., once in the positive direction, the original value of  $\theta$  is increased by  $2\pi$ . Crossing an internal boundary causes a jump of  $\pi$ .

The formulas and relations to be examined assume a simpler form if we replace the polar angle  $\theta$  by its quotient  $S$  by the angular frequency  $\omega$ , writing ( $S$  used in a new sense, later to be related to the earlier one):

$$(5.4) \quad \theta = \omega S.$$

Then equations (5.3) are replaced by

$$(5.5) \quad A = R \cos \omega S, \quad B = R \sin \omega S.$$

$S$ , like  $\theta$ , is multiple valued, being determined only to integral multiples of  $2\pi/\omega = 1/\nu$ , the period of vibration of the travelling wave.

The substitution of the expressions (5.5) into (4.2) gives, after equating coefficients of sines and cosines (as justified by the usual reasoning), the polar form of the Helmholtz equations:

$$(5.6) \quad |\nabla S|^2 - \frac{1}{c^2} = \frac{1}{\omega^2} \frac{\nabla \cdot (\rho \nabla R)}{R \rho}$$

$$(5.7) \quad \nabla \cdot (\rho R^2 \nabla S) = \nabla^2 S + \nabla S \cdot (2 \nabla \log R + \nabla \log \rho) = 0$$

The corresponding substitution into (4.3) and (4.4) give

$$(5.8) \quad E' = \frac{\omega^2}{4} R^2 (|\nabla S|^2 + \frac{1}{c^2} + \frac{1}{\omega^2} |\nabla \log R|^2) \rho$$

$$(5.9) \quad \vec{F}' = \frac{\omega^2}{2} R^2 \nabla S \rho.$$

Finally, the expression for our elementary travelling wave becomes

$$(5.10) \quad \psi: (\psi', \psi'') = (R \cos \omega(S - t), R \sin \omega(S - t)),$$

(replacing  $-t$  by  $+t$  for power flow opposite to  $\nabla S$ ).

Equation (5.9) shows that the flow-lines defined by (5.1) are in the direction of the gradient of the function  $S$ , and thus are normal to each surface  $S = \text{const.}$  Moreover, (5.10) shows that this family of surfaces, regarded as the loci of the equation  $S = t$ , can be described as surfaces moving with increase of time, and representing loci of constant phase--all exactly as described in Section 1. But there is one difference. If we consider two near-by surfaces,  $S = t$  and  $S = t + \Delta t$ , where  $\Delta t$  is small, the second will cut a length  $\Delta N$  off the normal to the first surface (drawn at some given point  $(x, y, z)$  on the latter); and  $\Delta N / \Delta t$  approaches, as  $\Delta t \rightarrow 0$ , the rate of motion of the surface of given phase,  $S = t$  (at that point). But this is easily seen to be the reciprocal of the gradient of  $S$ ; in other words, this rate of motion  $\Delta N / \Delta t \rightarrow 1 / |\nabla S|$ . Now equation (5.6) shows that this is equal to the sound speed  $c$  only if the right-hand member of this equation can be neglected; otherwise the phase velocity of the travelling wave will depend (through the function  $R$ ) on the wave and on its frequency  $\omega$ .

At this point the conventional treatments regard the right-hand member of (5.6) as approaching zero as  $\omega$  increases indefinitely. As a deduction from (5.6) and what preceded it, this is incorrect, since the other quantities,  $S$  and  $R$ , also depend on  $\omega$  (otherwise they could not satisfy Helmholtz' equation). It is perfectly possible for  $\nabla^2 R / R$  to become infinite as  $\omega$  increases, in such a way that the right hand member of (5.6) does not approach zero. In fact, as we shall see in Appendix D, counter-examples to the approach to zero can be given.

More logically, we could select, from the infinitude of waves satisfying the boundary conditions at the physical and internal

boundaries, a sub-infinitude for which

$$(5.11) \quad \lim_{\omega \rightarrow \infty} \nabla \cdot (\rho \nabla R) / R \rho \omega^2 = 0.$$

Many authors derive important local consequences from this assumption (cf. Hadamard<sup>3</sup>, "Lectures ..." or Courant & Hilbert<sup>3</sup>, Vol. II). A general feature of certain methods is to start with a solution  $S_0$  of the eikonal equation, the limiting form of (5.6) under assumption (5.11); and then to obtain special solutions of (3.1), or of (4.2) or its equivalents (5.6) and (5.7), by methods of successive approximations or in formal power series in  $1/\omega$ . The fact that only local results can be obtained in the general cases of variable  $c$  needed in long range propagation results from the nature of the multiple-valuedness of  $S_0$ , whose branch loci are surfaces in space or curves in the plane--as we have noted in Section 2 with regard to caustics. See Appendix D.

It is possible to select a sub-infinitude of travelling waves, parameterized by the frequency  $\omega$ , in such a way that the power flow lines approach an appropriate subset of rays: such a sub-infinitude will be defined in Section 6, and will be termed a "Fermat family" of travelling waves, since, as will be shown, they are the ones in which the flow of power is governed at high frequencies by the usual application of Fermat's Principle. The fact that this is not automatically the case--that many simple and physically natural families of travelling waves exist in which the power flow curves are not remotely represented by rays even as  $\omega \rightarrow \infty$ --is shown by the classical examples in Appendix D.

The complex notation. Before leaving the present subject, we re-cast some of our earlier expressions in the simpler and more usual complex notation--which we have deferred up to now, in order to make the physics and the logic more visible. As noted at the close of Section 4, we may write for our travelling wave pair, A,B, the complex function  $\phi = C = A + iB = \text{Re}^{i\omega S}$ , which of course satisfies Helmholtz' equation. The corresponding complex wave  $\psi$ , which satisfies D'Alembert's wave equation, is given by  $\psi = \phi e^{\pm i\omega t}$ , the + for a retreating, the - for an advancing wave (in the direction of  $\nabla S$ ). Thus (5.10) is replaced by

$$(5.12) \quad \psi = \text{Re}^{i\omega(S \mp t)}$$

while the density and flux of energy and their means are given as follows, the bar denoting the complex conjugate:

$$(5.13) \quad \begin{aligned} E &= (\nabla \psi \cdot \nabla \bar{\psi} + n^2 \psi_t \bar{\psi}_t) \rho / 2, \\ \vec{F} &= -(\psi_t \nabla \bar{\psi} + \bar{\psi}_t \nabla \psi) \rho / 2 \end{aligned}$$

$$E' = (\nabla \phi \cdot \nabla \bar{\phi} + n^2 \omega^2 \phi \bar{\phi}) / \rho / 4,$$

$$\vec{F}' = (\bar{\phi} \nabla \phi - \phi \nabla \bar{\phi}) \rho \omega / 4i$$

Note: A Common Fallacy. In spite of valid results derived by some authors from (5.11), in many elementary presentations of the subject, this equation--which is obviously a necessary condition for the validity of the ray approximation--is assumed to be sufficient. That this point of view cannot be held is shown by a counter-example in Appendix D, which gives a family of travelling waves satisfying (5.11) but not leading to the ray approximation. This requires us to give the detailed discussion of Section 6.

## 6. Asymptotic and Fermat Families of Travelling Waves

Statements concerning the behavior of travelling waves at high frequencies, such as the assumption that "at high frequencies the mean flow of power is guided by the rays", must, if they are to be given precise meaning and mathematical proof, imply that an infinite set of waves is under discussion--not individual ones--and that their limiting properties are being considered as their frequencies increase without limit. This means that we have to deal with a family  $\{\phi_\omega\}$  of travelling waves, each individual member  $\phi_\omega$  of which is given by (5.10), or more conveniently, by (5.12). We wish to explore the conditions under which, as  $\omega \rightarrow \infty$ , the lines of mean power flow  $\vec{F}'$  (the " $\Lambda$ -lines" as we shall term them) approach the appropriate rays (the " $L$ -lines") "associated" with them by mutual tangency at a given fixed point: the least that it would imply is that the unique  $\Lambda$ -line through each fixed (non-specialized) point  $P$  have a tangent whose direction vector (the unit vector  $\vec{F}'_\omega / |\vec{F}'_\omega|$ ) approaches a limit direction as  $\omega \rightarrow \infty$ .

The many possibilities presented by actual  $\omega$ -parameter families of travelling waves are illustrated by the six examples given in Appendix D, each of which is an elementary and classical case having a simple physical meaning. The first four examples deal with homogenous media (constant  $n, \rho$ ) so that the rays are all straight lines. The fifth is among the earliest studied problems of reflection and refraction across a plane boundary between media of different but constant  $(n, \rho)$ .

The first example is the square membrane clamped at its edges. There are no internal boundaries (curves where  $\phi=0$ ); but, for each  $\omega$ , a set of singular loci (points) around which the  $\Lambda$ -lines wind; as  $\omega \rightarrow \infty$  the singular points increase in number and density throughout the square: every one of its points is either a singular point or the limit of such points as  $\omega \rightarrow \infty$ ; and the  $\Lambda$ -lines have their curvatures increasing without limit. Obviously no ray treatment is possible in this case.

The fifth example is the classical case of reflection and refraction of plane waves at a plane interface: no internal boundaries, but a singular locus  $\phi = 0$  consisting of parallel equi-spaced lines on the interface, crowding indefinitely as  $\omega \rightarrow \infty$ ; no limiting direction of  $\Lambda$ -lines.

The second example is again a membrane, but clamped about its circular boundary. The center is the sole singular point for all frequencies; but the interior boundaries (curves where  $\phi=0$ ) are concentric circles, spaced according to the roots of Bessel functions, and with a spacing approaching zero as  $\omega \rightarrow \infty$ . The mean power flow or  $\Lambda$  lines are these circles and others concentric with them. Therefore our necessary condition--the approach of limits in direction (and position) of all  $\Lambda$ -lines--is satisfied. In spite of this, these lines have nothing to do with rays, which are all straight lines. This should dispel the glib generality that at high enough frequencies, the rays guide the mean (or other) flow of power!

The third and fourth examples are of radiation emitted from a point, in a membrane or in space, respectively, and thought of as extending without limit. In these examples both the  $\Lambda$ - and the  $L$ -lines are radial straight lines through the point radiator: in every associated pair  $\Lambda = L$  at all frequencies.

Appendix D follows these familiar cases with a counter-example to the notion that (5.11) is sufficient for the applicability of rays as  $\omega \rightarrow \infty$ : a family of plane waves satisfying (5.11) but having no limiting direction. Clearly more than (5.11) must be assumed. We start by giving the preliminary

P. DEFINITION I. By an asymptotic family  $\{\phi_\omega\}$  of travelling waves shall be meant one having the following property: Given any point  $P_0$  in the medium, not belonging to an exceptional figure  $\Gamma$  which is independent of  $\omega$  and of dimension at least 2 lower than the medium, there exists a neighborhood  $N_0$  of  $P_0$ , within which the directions of all  $\Lambda$ -lines approach limits, in a "regular and smooth" manner. More precisely, such that if  $P$  is any point fixed in  $N_0$ , the direction vector tangent at  $P$  to the unique  $\Lambda$ -line through this point (for sufficiently large  $\omega$ ) will approach a limiting vector direction as  $\omega \rightarrow \infty$ . Further, this approach will also hold for the space derivatives of this field of directions in  $N_0$  of orders up to and including the fourth. Finally, the convergence is in all cases uniform on  $N_0$ .

This somewhat long-winded definition can be shortened and made more elegant by assuming the analyticity of all  $\phi_\omega$ --as we may, in view of the considerations cited in Appendix C. Then all that is needed is the requirement of uniform approach to limits of the directions themselves



(i.e., of the three analytic functions contained in the symbol  $\vec{F}'_\omega / |F'_\omega|$ ) on some fixed 6-dimensional neighborhood  $N^6_0$  of  $P_0: (x_0, y_0, z_0)$ : the uniform convergence of their derivatives of all orders automatically follows. We may add that in all examples given in terms of special functions--as those of Appendix D--this situation always occurs when the  $\Lambda$ -lines do approach limits (as in the last three examples cited).

Consider the field of limiting directions of the  $\Lambda$ -lines: by the assumptions of differentiability and uniformity of approach, it is a "smooth" (i.e., 4th order differentiable--usually analytic) field, except on  $\Gamma$ . A first application of the classical theory of ordinary differential equations shows that through each point  $P_0$  not on  $\Gamma$  passes one and only one integral curve  $\Lambda_\infty$  (i.e., tangent at each point to the direction of the field at that point). A second application of the general theory shows that the  $\Lambda$ -curve through any  $P$  (not on  $\Gamma$ ) approaches the integral curve  $\Lambda_\infty$  through  $P$  of the limiting field, as  $\omega \rightarrow \infty$ , and that this approach is uniform (through the fourth order of differentiation at least) in any bounded region, however large, not containing points on  $\Gamma$ . [A most convenient form of the general theorems used here is in G.D. Birkhoff's "Dynamical Systems", Chapter II].<sup>6</sup>

We wish to show that the limiting curves  $\Lambda_\infty$  in the asymptotic family  $\{\phi_\omega\}$  are normal to a family of surfaces,  $\Sigma(x, y, z) = \text{constant}$ ; i.e., that a function  $\Sigma$  exists, "smooth" to at least the 4th order defined (but usually multiple-valued) except on  $\Gamma$ . Here we apply a theorem of advanced calculus\* that states that a field of directions, described, e.g., by a unit vector  $\vec{U} = \vec{U}(x, y, z)$  (whose components have the present order of smoothness) is a field of normals to a family of

---

\*References in Appendix D.

surfaces of the above type if and only if the field is perpendicular to its curl; i.e.,  $\vec{U} \cdot \nabla \times \vec{U} = 0$ . This property will be established, in view of the way we have assumed the field of directions  $\vec{U}_\omega = \vec{F}'_\omega / |F'_\omega|$  converges to  $\vec{U}$ , once it has been proved for the latter. To do this, we first consider any point  $P_0$  at which  $\phi_\omega \neq 0$  and take a neighborhood  $N_0$  throughout which  $\phi_\omega \neq 0$ : in  $N_0$  the functions  $\Theta$  and  $S$  are defined, e.g. by (5.3) and (5.4); and  $\vec{F}'_\omega$  is normal to the surface  $S = \text{constant}$ , (5.9). Hence, this vector field, and therefore  $\vec{U}_\omega = \vec{F}'_\omega / |F'_\omega|$ , is perpendicular to its curl. Now, as shown in Appendix C, even on internal boundaries the directions  $\vec{N}_\omega$  are defined, and join smoothly onto those defined at other points. Hence the property of being perpendicular to its curl is extended by continuity to the internal boundaries: it therefore holds for the limiting vector field  $\vec{U}$  at every point not on  $\Gamma$ , whose dimension is, according to the above definition, at least two units lower than that of the medium. This suffices to show that the property of the  $\Lambda$ -lines of being normal to a family of surfaces ( $S = \text{constant}$ , one and only one through each point) is transmitted to their limits, the L-lines.

It is appropriate to emphasize the purely geometrical nature of the results established above for asymptotic families: from their definition in terms of limiting directions at each point (not on  $\Gamma$ ) we have concluded that the  $\Lambda_\omega$ -line through each point (not on  $\Gamma$ ) approaches a limit curve  $\Lambda_\omega$  through that point; and further, that the very special property of orthogonality to each of a family of surfaces--one and only one of which passes through each given point not on  $\Gamma$ --applies also to the limit curves. While the proof outlined above is based on analytic

formulas, the result may be regarded as geometrically evident. So also may be the fact that the unique surface ( $\Sigma = \text{const.}$ ) through any given  $P$  (not on  $\Gamma$ ) normal to the limiting  $\Lambda_\infty$ -lines is actually the limit, as  $\omega \rightarrow \infty$ , of the surfaces ( $S_\omega = \text{const.}$ , etc.) through  $P$ . An outline of a deeper discussion of these matters will be given in Appendix D.

It is also appropriate to note that these geometrical relations (or their analytical proofs) do not automatically establish the fact that the functions  $S \rightarrow \Sigma$  as  $\omega \rightarrow \infty$ . This is because a family of surfaces, such as  $\Sigma = \text{const.}$ , does not determine the function  $\Sigma$  that may be used to represent it: thus  $\Sigma^3$  could serve to represent the above family as well as  $\Sigma$ ; and, in fact, any such family can always be represented in infinitely many different ways; this however would imply that the new function is a function of the first--of the form  $f(\Sigma)$ . Since obviously  $\nabla f(\Sigma)/|\nabla f(\Sigma)| = \nabla \Sigma/|\nabla \Sigma|$ , the normals define the same direction fields. This state of affairs is illustrated by the fact that in the second example of Appendix D,  $S_\omega = p\theta/\omega$  ( $\theta = \tan^{-1} y/x$ , the polar coordinate angle, while the integer  $p$  is independent of  $\omega$ ). Hence  $S_\omega \rightarrow 0$ , whereas  $\Sigma \neq 0$ , and can be taken, e.g., as  $\theta$  or  $p\theta$ .

As noted before, in view of this same example, the limiting  $\Lambda_\infty$  through a point  $P$  not on  $\Gamma$  may or may not be the ray  $L$  associated with it by mutual tangency at  $P$ : to justify any application of the ray method, conditions that  $\Lambda_\infty = L$  must be established. An obvious criterion is that the function  $\Sigma$  in the representation ( $\Sigma = \text{const.}$ ) of the family of surfaces normal to the  $\Lambda_\infty$ -lines may be so chosen as to satisfy the eikonal equation  $|\nabla \Sigma|^2 = n^2$ . Since, as we have observed, the whole

class  $f(\Sigma)$  represent this same family of surfaces--and if one satisfies the eikonal the others in general will not--the application of this criterion is somewhat impracticable. Another criterion reposes on the intrinsic geometrical shape of the  $\Lambda$ -and  $L$ -lines associated at the point  $P$ : their curvature and (in the case of space curves) their torsion. Appendix D, after recalling the definitions of these fundamental concepts (and giving references), establishes formulas for them in the cases of associated  $\Lambda$ -and  $L$ -lines. Now clearly if these two curves are to coalesce as  $\omega \rightarrow \infty$ , the curvature and the torsion of these two curves must differ by an amount approaching zero. The formulas then show that our asymptotic family  $\{\phi_\omega\}$  must be such that not only must (5.11) be valid in the sense of uniform convergence in some neighborhood  $N$  of each point  $P$  (not on  $\Gamma$ ), but the derivatives of the left-hand member of (5.11) up to the second order must converge to zero in the same way.

(5H) The augmented version of (5.11) italicized above shall be referred to as (5H). So far it has been established only as a necessary condition for the ray treatment of power flow in an asymptotic family. We now have the following

THEOREM I. Condition (5H) is not only necessary but sufficient for the validity of the ray treatment of power flow in an asymptotic family of travelling waves  $\{\phi_\omega\}$  at high frequencies.

Equation (5.6) combined with (5H) shows that  $|\nabla S_\omega| \rightarrow n$  as  $\omega \rightarrow \infty$ . On the other hand, the definition of asymptotic families states that the direction vector  $\vec{U}_\omega = \nabla S_\omega / |\nabla S_\omega|$  approaches the limit  $\vec{U}_\infty$ . Therefore the

product  $\nabla S_\omega = |\nabla S_\omega| \vec{U}_\omega$  approaches the product of the limits,  $n \vec{U}_\infty$ . Finally, since  $S_\omega = \int \nabla S_\omega \cdot d\vec{r} + \text{const.}$  ( $\vec{r}$  is the position vector), the integration being along any fixed regular path, it follows (by the uniformity of the convergence of the integrand) that, as  $\omega \rightarrow \infty$ ,  $S_\omega$  approaches a limit  $S_\infty = \int n \vec{U}_\infty \cdot d\vec{r} + \text{const.}$  and that  $\nabla S_\infty = n \vec{U}_\infty$ , so that, on taking  $\Sigma = S_\infty$ , we can satisfy the eikonal equation, since  $|\nabla S_\infty|^2 = n^2$ . From this point, the elementary general theory (if Osgood, Chapter XIV.23)<sup>4</sup> shows that the limits of the  $\Lambda$ -lines are in fact a family of rays. All this assumes that we are avoiding points of  $\Gamma$ . We pass to the next preliminary

P. DEFINITION II. By a Fermat family of travelling waves shall be meant an asymptotic family  $\{\phi_\omega\}$  satisfying the condition (5H)--or, equivalently, the condition that  $S_\infty = \lim S_\omega$  exists and satisfies the eikonal equation.

Here we may observe that the reason for introducing  $S = \omega\theta$  in Section 5 instead of the apparently more natural  $\theta$  is this simpler behavior of  $S$  in Fermat families. In the second example in Appendix D--the asymptotic but not Fermat family--it is  $\theta$  and not  $S$  that approached the required limit.

The rest of the present study will be based on the notion of Fermat families--a concept to which we have been led in our attempt to justify the ordinary ray treatment. But certain questions of crucial importance introduce themselves: first, can we satisfy the physically realistic boundary conditions imposed on the problems of hydro-acoustic propagation by a Fermat family, or do we have to use linear combinations

of such families? and second, in the latter case, is such a combination of Fermat families also a Fermat family? And other issues can be raised.

The answer to the first question will depend on the precise nature of the boundary conditions. If they are of the classical deterministic type, they cannot in any but the simplest cases (e.g., in the third and fourth examples of Appendix D) be satisfied by a Fermat family. Even as simple a case as the fifth example, of reflection and refraction of plane waves at a plane boundary--with a resulting non-asymptotic family of travelling waves--shows the impossibility mentioned. The same examples, which combines Fermat families linearly, i.e., by the "principle of superposition" shows that the second question has a negative answer. In fact we have the following:

THEOREM II. If  $\{\phi_\omega\}$  and  $\{\phi'_\omega\}$  are two linearly independent asymptotic families of travelling waves, their combination by the principle of superposition, i.e.,  $\{a\phi_\omega + b\phi'_\omega\}$ , is never an asymptotic family (unless  $a = 0$  or  $b = 0$  or they tend to coalescence in the limit).

The proof is given by a direct calculation based on the expressions (5.12) and (5.13).

Of course the classical boundary conditions and the combination by the principle of superposition would also automatically prove that multi-path transmission is impossible. Since we know experimentally that it does occur and is very important, we are driven to the conclusion that boundary conditions formulated as implied above do not represent the physical facts in such cases: that we have used the "wrong

model". The following theorem shows one way out of this reflection problem:

THEOREM III. If in the combination of the two asymptotic families in Theorem II, we insert a "relative phase factor", i.e., replace  $b$  by  $b e^{i\delta}$  (where  $\delta$  is real), then calculate the components of  $\vec{F}$  for this linear combination by formulas (5.13) and lastly take the mean of the result over a full period (i.e., integrate with respect to  $\delta$  between 0 and  $2\pi/\omega$  and divide the result by this period), it is found that the components of the flux vector  $\rightarrow$  the linear combinations, by the real factors  $|a|^2$  and  $|b|^2$ , of those of the original wave families.

By letting  $\omega \rightarrow \infty$ , this linear combination of these important physical quantities approaches the corresponding combination of their limits: indirectly, then, we can apply the asymptotic or the Fermat families--and thus restore the basis for the validity of ray methods.

What is the physical basis of the phase averaging of Theorem III? One answer (applicable to specular reflection) would be that the actual position of the reflecting plane is indeterminate within the limits of very short wave lengths, and thus produces the incoherence assumed. But a more general answer is to be found in the medium itself. These matters will be examined further in Sections 10 and 11--most conveniently after the Hamiltonian picture of ray phenomena has been introduced, in Sections 7-9.

This extension of power flux to phase space will be based in Sections 10 and 11 on the geometrical ideas just developed, for the sake of intuitive simplicity. More sophisticated methods would do the corresponding task for the broadened conceptions to which we now turn.

Broadening of the definitions. It is necessary to broaden the two preliminary Definitions I (asymptotic families) and II (Fermat families), to apply to the case of multi-path transmission, as when two or more rays pass through the points of emission and reception, and when this does not involve the type of reflection at a surface of discontinuity noted above, but continuous refractive bending. Then the ray directions throughout extensive volumes of the medium will not be unique: at any point P there will be as many different directions as there are rays joining it with the emitter. Since, on the other hand, for each fixed  $\omega$ , however large, the energy flow directions given by (5.13) at P are uniquely determined (when P is not on the 2-lower dimensional  $\Gamma$ ), it follows that the latter single direction at P could not possibly approach the several ray directions through this point: no asymptotic family according to our preliminary definition could exist at such ranges from the emitter as allow multi-path transmission. In physical terms, such transmission will produce a complicated interference pattern, varying rapidly with increasing  $\omega$ , and preventing the approach to any limit.

This interference behavior results, of course, from the time differences of arrival at P of an emitted phase as it takes the several paths: the resultant at P will depend strongly on the frequency  $\omega$ . Now this fact suggests the way out of our difficulties: carry the whole picture from that in the space occupied by the medium (xyz-space) to a picture in the "space-time" of the variables  $x, y, z, t$ . For in all cases of radiant energy in classical physics, the path ("world line") connecting the point ("event")  $(x^0, y^0, z^0, t^0)$  with  $(x, y, z, t)$  is unique (with rare exceptions



which cause no difficulties)\*. In other words, the times taken for a phase to travel from  $(x^0, y^0, z^0)$  to  $(x, y, z)$  are different for the different rays connecting these points, so that  $t - t_0$  will determine the ray in question.

This transference to space-time requires us to return to the time dependent elementary travelling waves (5.12)  $\psi = \text{Re}^{i\omega(S-t)}$ , and the corresponding  $E$  and  $\vec{F}$  given by the first two equations in (5.13). Then in rephrasing Definition I we should have to consider limits at a fixed point-and-time ("event")  $(x, y, z, t)$  of these and related quantities as  $\omega \rightarrow \infty$ . We also have to consider, not simply the 3-vector  $\vec{F}$  but the "4-vector"  $(\vec{F}, E)$ . Accordingly, we broaden the earlier definitions as follows:

DEFINITION I. An asymptotic family  $\{\psi_\omega\}$  of elementary time-dependent travelling waves is one which in a neighborhood of each point-and-epoch  $(x, y, z, t)$  has its 4-vector  $(\vec{F}, E)$  approaching a limiting 4-direction - i.e., the ratios of its four components approach limits.

Here it is understood that points on a 2-lower dimensional  $T$  may be excepted; and that the convergence is uniform in the neighborhood, extended slightly to complex values of the variables.

In order to reach a corresponding definition of the Fermat family, it is useful to introduce the velocity vector  $\vec{V}$  of flow of energy, regarded as located in the medium. Clearly  $\vec{V} = \vec{F}/E$ , and with an asymptotic family, this approaches a limit as  $\omega \rightarrow \infty$  (with the usual exclusions).

Locally, (5.12) and (5.13) show that (5.8) and (5.9) apply to  $\vec{F}$  and  $E$ ,

\*Being, as explained in Appendix G, confined to 2-dimensional manifolds (the 1-dimensional caustics in the  $(x, y, t)$  case).

but with the second members multiplied by 2. The length  $V$  of  $\vec{V}$  is  $V = F/E$  ( $F = \text{length of } \vec{F}$ ). Our equations, after taking reciprocals involving  $1/V$ , lead to the following:

$$(16.1) \quad \left( |\nabla S| - \frac{1}{V} \right)^2 + \left( \frac{1}{c^2} - \frac{1}{V^2} \right) + \frac{1}{\omega^2} |\nabla \log R|^2 = 0.$$

The first conclusion is that  $V \leq c$ , since otherwise the second parenthesis would be positive, and since the other two are non-negative, the sum could not vanish. The second conclusion is that  $V$ , which depends on  $\omega$ , will approach  $c$  if and only if  $|\nabla \log R|/\omega \rightarrow 0$ . This is again based on the non-negativeness of the first and second parenthesis. Turning to (5.6) we see that the above approach will occur if and only if (5.11) is valid - and with it, the consequence that  $|\nabla S| - 1/c \rightarrow 0$ . These results, as noted earlier in this section, establish the approach to coincidence of the  $\Lambda$ -line through a given  $(x, y, z, t)$  to the  $L$ -line tangent to it at that point, as  $\omega \rightarrow \infty$ . Thus our definitive:

DEFINITION II. A Fermat family is an asymptotic family in which the limiting 4-directions (ratios of the four components of  $(\vec{F}, E)$ ) are such that the energy speed  $V$  is the phase speed  $c$ .

Our previous examples show that many families  $\{\psi_\omega\}$  can fail to be asymptotic; that many asymptotic families exist which are not Fermat; and that Fermat families exist. Furthermore, Theorems I and II continue to be valid with the present broadened definitions of these terms.

It remains to show that under the physical conditions of importance to long range propagation, with smoothly varying but quite general physical functions  $(\rho, c)$  Fermat Families exist. This task will be carried out in Appendix D; but it is appropriate to close the present section with three remarks.

First, the literature in which mathematically correct validations of ray treatments on the basis of the wave equation, while very voluminous, does not appear to cover the case of multi-path transmission of acoustic energy, with caustics causing complicated multiple-valuedness of the phase function. Thus in Courant & Hilbert<sup>3</sup>, Vol. II, Chap. VI, Sec. 5, the equations (3) and (20) contain such functions as " $\psi$ " (our S) which at ranges of interest to us become multiple-valued. See also the references given in that place.

Second, the method that we shall use in Appendix G is based on the elementary solution first established in sufficient generality by J. Hadamard<sup>3</sup> in 1923 (Lectures on Cauchy's Problem, etc.) In our case of time-independent coefficients, slowly varying in range, this solution simplifies and can be extended to the long ranges required. Furthermore, our assumptions of analytic data allow great simplifications: by contour integration in the complex domain of the initial time of a harmonic factor times the elementary solution, we obtain our Fermat families of emitted travelling waves without the need of Hadamard's complicated "finite part" of a divergent integral: or the even greater complications of the "improper functions" (L. Schwartz' "Distributions", including in particular Dirac's Delta Function) which are common methods in the present mathematical

theory, used largely to avoid assumptions of analyticity and use of complex independent variables: to economize assumptions for mathematical form rather than physical necessity.

Third, our proof in Appendix D requires a physically unimportant but mathematically crucial broadening of the notion of limit in our above Definitions I and II: instead of "strong limits" (actual approach at each point  $(x,y,z,t)$  — with local uniformity) we must use "weak limits": the approach of integrals of our functions over arbitrarily small and given regions (not changing with  $\omega$ ). After all, it is not the density  $\vec{F}$  itself, but the total flux of energy it produces across a small but finite piece of surface during a small finite interval of time that is physically meaningful and measurable. Similarly for the integral of  $E$  over a small volume at a given instant (epoch).

## 7. A Mechanical Analogue Model of Acoustic Rays

Having established the close coincidence of rays with the lines of power flow at high frequency, our next step is to establish a better way of applying rays to transmission loss problems than the usual one, outlined in Sections 1 and 2. For this purpose it is necessary to borrow from the classical theory of dynamical systems certain mathematical tools. The exact and general approach will be set forth in the succeeding sections. The present section is intended to introduce the ray versus particle-trajectory ideas in a simple form that can be visualized. A still more elementary account of the particle-wave correspondence (in very special cases) is given in "Physics" (Physical Science Study Committee, D.C. Heath & Company, Boston 1960): Chapter 15 (The Particle Model of Light) studies surface refraction by essentially the model used in this section.

One purpose of the mechanical model described herein is to aid in bringing mechanical intuition to bear on the behavior of sound propagation by rays in a vertical plane. Such propagation occurs when the sound speed  $c$  depends not only on depth  $z$  but also on horizontal distance  $x$  along the direction of propagation--but not on distance across it; thus  $c = c(x, z)$ . The special case when the acoustic profile is the same at all points, is included in the model; this case, which is studied in Appendix B, is very simple, and the rays are easy to visualize, being periodic (see Figure B1). As shown in Appendix B, they are given mathematically by explicitly solving their differential equations by quadratures. The construction and operation of the model

will be described here, its physical justification being deferred to Section 8.

The construction and validity of the model are based on two facts. Firstly, as will be shown in Section 8, the equations determining the acoustic rays are identical with those governing certain motions of a particle of unit mass under the sole action of a field of force derived from a potential  $V = -C^2/2c^2$ ; here the constant of proportionality  $C$ , of dimensions of velocity squared, is chosen arbitrarily and determines the scale of the construction. Secondly, in the two-dimensional case of rays in the  $xz$  - plane, the motion under the action of the potential  $V = V(x,z)$  is equivalent (to quantities of higher order) to the motion of a unit particle constrained to lie on an appropriately defined smooth surface  $\Sigma$  and acted on only by gravity and the normal reaction of  $\Sigma$ .

The construction of  $\Sigma$  is as follows: First, turn the  $xz$  - plane from its originally vertical position to a horizontal one; second, introduce a third axis,  $y$ , directed downward; finally, construct  $\Sigma$  as the locus of the equation

$$y = -V/g = C^2/2c^2g$$

where  $g$  is the constant of gravity; thus  $V = -gy$ , the gravitational potential.

If, then, the particle is started at any point  $P_0$  of  $\Sigma$  in any direction (tangent to  $\Sigma$ ) and with the speed  $v_0 = C/c_0$  ( $c_0 = c$  at  $P_0$ ), its path will (to quantities of higher order in the vertical component of its velocity) trace out the ray tangent to the same initial direction. Figure 7.1 illustrates the situation in the case

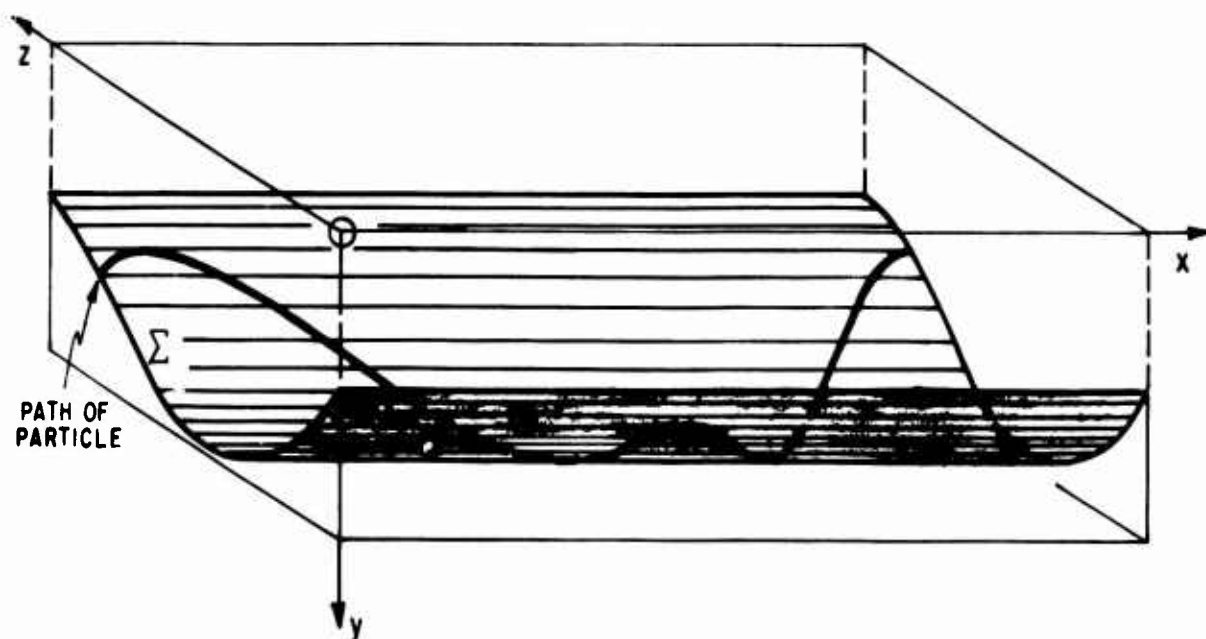


FIGURE 7.I CASE  $c = c(z)$ :  $\Sigma$  CYLINDRICAL

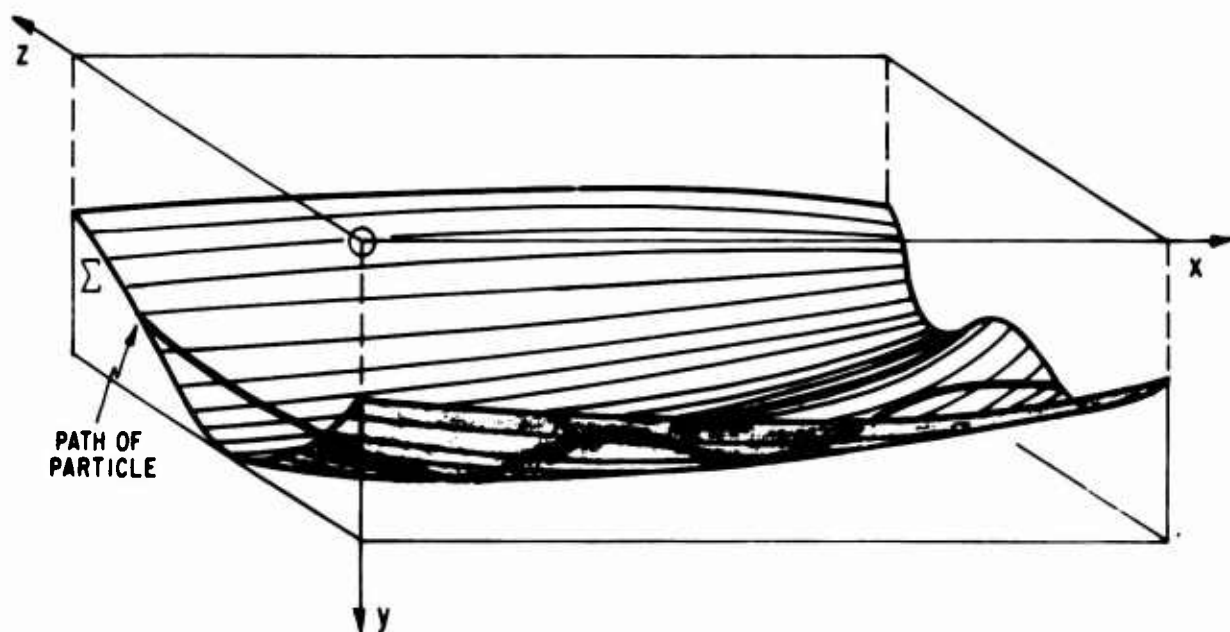


FIGURE 7.II CASE  $c = c(x, z)$ , WITH DUCT DOUBLING

of  $c = c(z)$  and the motion gives periodic rays; Figure 7.5 shows the more general case  $c = c(x, z)$ .

The above choice of initial speed  $v = v_0$  makes the total energy  $E = v^2/2 + V = 0$  initially--and hence all along the path--a requirement that will become clear in Section 8. Therefore, along each path,  $v^2 = 2gy$ , so that  $v$  has its minimum at the highest point reached on  $\Sigma$  by the ray. The equation defining  $\Sigma$  shows that it is always well below the  $xz$ -plane. Hence  $v$  is never zero. The only exceptions may be at the boundaries of the medium: If we wish our model to illustrate bottom absorption, we note that the bottom is a vertical plane or cylindrical surface in Figures 7.1 and 7.2 on which  $z$  has relatively large values. Then we must think of it as made out of an absorbing sticky putty. If we wish to represent surface reflection, we must bound  $\Sigma$  by a plane parallel to the  $xy$ -plane with a relatively small constant  $z$ , and assume it is made of a perfectly elastic material: at impact, the velocity vector will suffer a discontinuity, but without resulting change in speed.

A bundle of power-transmitting rays can be modelled as a set of many mutually non-interfering particles, initially projected from a point (the "emitter"). Their rarefactions and condensations (at caustics) as well as their splitting into ducts (Figure 7.3) can be visualized.

While the pictures of the model shown in Figures 7.1 and 7.2 are schematic and give a qualitative idea of the process of rays traced by the moving particle, a further step seems desirable, namely obtaining actual dimensions in representative cases. For this purpose, we must first fix our units and next decide on a convenient value of the general scale factor  $C$ . We shall use the MKS system (meter-kilogram-second).



On the basis of extensive oceanographic observations, the sound speed in most cases lies in the interval between 1475 M/S and 1525 M/S, only very rarely attaining either extreme. We shall, accordingly, measure  $c$  about the central value  $c_1 = 1500$  M/S, setting  $c = c_1 + u$ : this "algebraic excess" speed  $u$  lies between  $\pm 25$  M/S.

As for  $C$ , a convenient scale for our model is obtained in two steps: First, we set  $C = c_1 \sqrt{2gh}$ , where  $h$  is a length--the linear scale factor (it is seen that this gives  $C$  the dimensions of a velocity squared, as required earlier). Next, after substituting this expression into the equation of  $\sum$ , we select that length  $h$  which gives the model a convenient size. We have

$$\begin{aligned}\sum: y &= \frac{C^2}{2gc^2} = h \frac{c_1^2}{c^2} = h \left(1 + \frac{u}{c_1}\right)^{-2} \\ &= h \left(1 - \frac{2u}{c_1} + \dots\right),\end{aligned}$$

the dots in the binomial expansion denoting terms of higher order in the small quantity  $u/c_1$ . This ratio lies between  $\pm 25/1500 = \pm 1/60$ , so that the second term in equation for  $\sum$  is between  $\pm 1/30$ .

To arrive at a useful choice of the scale constant  $h$ , we must realize that the deep channel propagation of interest here requires an ocean depth between two and three miles at least: let us take 3 kilometers as the length of the interval of values of  $z$  in our model ( $0 \leq z \leq 3 \times 10^3$  M). Further, the usual convergence zone distance, of the order of thirty miles (about 50 kilometers) requires the model to be about  $5 \times 10^5$  M long, so that some ten ray periods occur. Clearly, such

a  $3 \times 10^3$  M by  $5 \times 10^5$  M area can be viewed only from the altitude of an airplane. If the hills and valleys of the surface  $\Sigma$  are to be visible from such a distance, they must be of the order of hundreds of meters. Let us accordingly take  $h = 10^4$  M. Then the xz-plane in our figures is  $10^4$  M above the distance of the median horizontal plane of  $\Sigma$ , which lies between parallel planes  $333 \frac{1}{3}$  M above and below the median plane. Such changes in altitude  $y$  of  $\Sigma$  could easily be seen from our airplane, and the motion of the particles discerned. If it were night and they were luminous, they would leave a path on a fixed camera, which would thus perform a mechanical analogue ray-tracing.

To get the model to laboratory size, we can, without changing the mechanics of the model, scale it down by a factor of  $10^4$ . Then it becomes 30 cm wide, 50 M long (or less, if one does not require ten convergence zones), and the median plane 100 cm deep, the surface  $\Sigma$  lying between two planes 3.33 cm above and below the median. To indicate that the median is so much lower than the xz-plane, we have in Figures 7.I and 7.II joined these planes with dashed lines, indicating that their lengths are out of scale.

## 8. Rays, Particle Dynamics, and Hamiltonian Phase Space

The acoustic rays being, by Fermat's Principle, the "paths of least time" between two points, are the solutions of a certain system of ordinary differential equations. These equations result from the application of the elements of the calculus of variations (Euler-Lagrange equations) to this time-minimizing problem. From the mathematics of these ray differential equations it follows that, through any given point and in any given direction, there passes one and only one ray. On the other hand, through two points, there are more possibilities: there may be no ray joining them (shadow zones); just one ray; or several (multi-path transmission). These facts and their analytic formulation are given in Appendix A, and developed in more detail in a special case in Appendix B. There, as everywhere else in the present study, the independence of sound speed on time is assumed.

It has been known since the early developments of dynamical theory that the rays in a refractive medium of generally variable index of refraction  $n = 1/c$  (i.e., heterogeneous but isotropic), and a certain class of trajectories of a particle of unit mass, subjected solely to forces deriving from a potential field  $V = -n^2/2 = -1/2c^2$ , are identical. There are two provisos for this ray-particle trajectory equivalence:

First, the time  $t$  in Fermat's Principle must be replaced in the particle's motion by a different variable, which we shall call the "pseudo-time" and denote by  $\tau$ . Their relationship is expressed as follows ( $s$  denoting arc length):

$$(8.1) \quad ds/dt = c ; \quad ds/d\tau = 1/c = n ; \quad d\tau = c^2 dt .$$

Along each given path or ray, these equations determine, by a simple integral, any two of the quantities (s, t,  $\tau$ ) in terms of the third. The fact that the greater the sound speed  $c = ds/dt$ , the smaller the particle speed  $ds/d\tau = n$ , so that the latter moves faster in a "denser" medium (higher index  $n$ ) was most troubling to the Newton-Huygens "corpuscle" versus "wave" arguments about the nature of light. On examining Figures 7.I and 7.II (Section 7), we see that the particle moves faster (more kinetic energy) the lower its position on the surface  $\Sigma$  (less potential energy) i.e., the lower its sound speed).

Second, not all motions of the particle trace out rays--only those having a certain given value of the total energy. With the arbitrary additive constant in the potential energy chosen as we have done above, the appropriate value for the total energy must be zero. We recall that this was accomplished by the choice of initial speed of the particle in the model in Section 7. This second condition for equivalence of rays and particle paths may seem less surprising than the one of the preceding paragraph, since from a given initial point each initial velocity vector (with 3 independent components) determines one and only one particle trajectory, whereas the ray through that point is determined by its direction (2 independent parameters, e.g. direction angles). There are therefore 6 variables needed to specify a "state of motion" or "phase" (position & velocity; equivalently, position & momentum) of the particle, in contrast with 5 variables for the

position & direction phase for the rays. As we shall see, this lower dimensional space is the locus of the equation setting the total energy equal to zero.

The proof of the equivalence under the two hypotheses just stated is in three steps: the first is to write down the differential equations resulting from the application of Newton's laws of motion to the particle. The second step is to apply our first assumption, changing the independent variable  $\tau$  in the equations of motion to  $t$  by the use of (8.1); and compare the result with the ray differential equations, (A.2) of Appendix A.

In rectangular coordinates  $(x,y,z)$  the equations of motion of the unit particle are

$$(8.2) \quad \frac{d^2x}{d\tau^2} = - \frac{\partial V}{\partial x} = \frac{\partial}{\partial x} \frac{n^2}{2}, \dots$$

the dots standing for the corresponding equations in  $y$  and  $z$ . Along any given particle trajectory, (8.1) gives the following equations for change of variable of differentiation from  $\tau$  to  $t$ :

$$d/d\tau = d/c^2 dt = n^2 d/dt ;$$

hence

$$\frac{d^2x}{d\tau^2} = n^2 \frac{d}{dt} \left( n^2 \frac{dx}{dt} \right).$$

After inserting this into (8.2) an obvious manipulation yields the first ray equation (A.2), as desired; etc.

The last step brings into play our second assumption, which can be expressed either by the special choice of the additive constant in the potential energy term  $V$ , implied by setting  $V = n^2/2$ , or by requiring that the total energy along the sub-class of paths we are considering be zero: in view of (8.1), either assumption is equivalent to the other.

In the language of classical mechanics, we have derived Maupertuis' principle of least action from Newton's equations for our particle, and have then identified the "action" with the "time"  $t$  along each ray, so that Fermat's principle of least time is obtained. Mathematically, we have merely re-written differential equations in a different form--and re-verbalized them, so that one physical picture is replaced by another. The utility of this process is that it guides us in applying a highly developed branch of mathematics (modern dynamics) to our difficult problems of long-range transmission in a heterogenous medium, affected by random factors. Our first step is to write the differential equations of our problem in Hamilton's canonical form.

In addition to the coordinates  $x, y, z$  of our particle, we introduce three more variables, the components of its momentum,  $p_x, p_y, p_z$ . Then we introduce the following function  $H$  of these six variables, the Hamiltonian of our dynamical system

$$(8.3) \quad H = 1/2 \left( p_x^2 + p_y^2 + p_z^2 - n^2 \right).$$

Using the symbole  $\partial$  to denote partial differentiation with respect to the following set of seven independent variables (i.e., holding all but one fixed):

$$t, x, y, z, p_x, p_y, p_z,$$

and the symbol  $\frac{d}{d\tau}$  to denote differentiation along a given trajectory (so that only  $\tau$  is varied arbitrarily, the other variables being functions of it, as explained below), Hamilton's canonical equations for our dynamical problem are the following

$$\begin{aligned} \frac{dx}{d\tau} &= \frac{\partial H}{\partial p_x}, \quad \frac{dy}{d\tau} = \frac{\partial H}{\partial p_y}, \quad \frac{dz}{d\tau} = \frac{\partial H}{\partial p_z} \\ (8.4) \end{aligned}$$

$$\frac{dp_x}{d\tau} = -\frac{\partial H}{\partial x}, \quad \frac{dp_y}{d\tau} = -\frac{\partial H}{\partial y}, \quad \frac{dp_z}{d\tau} = -\frac{\partial H}{\partial z}$$

The first three show that the momentum vector is the velocity vector (with respect to the pseudo-time  $\tau$ ) of our unit particle, since they give  $dx/d\tau = p_x$ , etc. The last three equations become, on replacing the momenta by these values, identical with the Newtonian equations of motion. Therefore (8.4) and (8.2) are identical and we can use either system to determine the motions of the particle. What about the total energy? This is evidently equal to  $H$ . If any trajectory is given (i.e., any set of six functions  $(x, \dots, p_z)$  of  $\tau$  satisfying (8.4)), the rate of change of  $H$  along it is found by the rules of partial differentiation to be

$$\frac{dH}{d\tau} = \frac{\partial H}{\partial x} \frac{dx}{d\tau} + \dots + \frac{\partial H}{\partial p_x} \frac{dp_x}{d\tau} + \dots$$

which is seen at once from (8.4) to be zero. (Note that this would not have been true if  $c$  contained the time). Hence  $H = h$ , a constant: the "law of conservation of energy" for this system.

Now suppose we select the set of trajectories for which this constant  $h$  is zero. The first three equations in (8.4) give

$$p_x^2 + p_y^2 + p_z^2 = (dx/d\tau)^2 + (dy/d\tau)^2 + (dz/d\tau)^2 = (ds/d\tau)^2$$

Therefore, the assumption that  $H = 0$  gives, according to (8.3), the relation  $(ds/d\tau)^2 = 1/c^2$ , and hence  $ds/d\tau = \pm 1/c$ , the (+) if we use the convention of measuring  $s$  in the sense of increasing  $\tau$ , the (-) with the opposite convention. We shall use the former convention hereafter. Thus, we have established (8.1), and hence by the earlier reasoning, the equivalence of the sub-class of solutions of (8.4) for which  $H = 0$  with the rays--the solutions of (A.2).

We can now give a more geometrical meaning to the momenta along a ray. Applying (8.1) to (8.4) we see that

$$p_x = dx/d\tau = dx/cds = \cos x/c = n \cos x,$$

and similarly for the other components; thus we express the momenta as the index times the direction cosines:

$$(8.5) \quad p_x = n \cos \alpha, \quad p_y = n \cos \beta, \quad p_z = n \cos \gamma.$$

The direction angles ( $\alpha, \beta, \gamma$ ) are made with the positively directed coordinate axes by the half-tangent to the ray in the sense of increasing  $s$ .

Our results lead to the following picture: The six values of the coordinates-and-momenta of our particles constitute a possible phase for it; the totality of such phases--or "points" in "6-dimensional space"--is called phase space. The locus in this space of the equation



$H = 0$  is a 5-dimensional manifold, the set of all possible coordinates-and-directions on our acoustic rays. By the basic theorems of systems of differential equations of the first order, such as (8.4), through each point of phase space passes one and only one trajectory (path in phase space) of our particle; and if the point in phase space is located on the manifold  $H = 0$ , the unique trajectory through it will remain on this locus: the 5-dimensional manifold  $H = 0$  is "made up of" those phase-space paths which are the rays--or, in more familiar terms, whose projections onto the physical xyz-space of the medium are the rays. To the question: "How does a point on a ray in the latter space acquire two extra coordinates?" the answer is: "by having the two variables of the direction of its tangent recorded". It has "ascended" into the 5-dimensional space of the position-and-direction of the point which traces it: a "lift" is the modern mathematical term for this.

We may see even at this stage one feature of the mechanism by which the ambiguities at caustics are resolved. In the 2-dimensional  $(x,z)$  case shown in Figure 2.II two rays  $A_1A_6$  and  $B_1B_6$  that intersect at a point  $P$ , always have different directions at  $P$ , and hence correspond to different points of phase space (3-dimensional, in the case of propagation in a plane). As  $P$  approaches a point  $P_0$  on the caustic of Figure 2.D, the third coordinate (that of direction) of each of their corresponding points in phase space approaches a common value, the direction of the caustic at  $P_0$ : there is a continuous change of each ray in phase space into the unique ray  $C_1C_6$  tangent at  $P_0$  to the caustic. Otherwise stated, we can continuously change the ray  $A_1A_6$  into  $B_1B_6$  by

starting with  $A_1A_6$ , moving the point from  $P$  to  $P_0$ , and then back to  $P$ --but now using, in reverse, the sequence of directions of the lines through which  $B_1B_6$  passed in going into  $C_1C_6$ .

A very simple model (to be returned to in Section 16) shows the essential ideas; propagation in the vertical  $xz$ -plane is assumed. Let us start with a cylindrical surface, that we may think of as of circular cross section. On it a family of helices is traced, all having the same pitch, so that one and only one passes through each point. This could be constructed by ruling parallel sloping lines on a rectangular sheet of paper, and rolling it up into a cylinder (the radius being chosen so that the lines join along the edges). Now think of the cylindrical surface as transparent and the lines opaque; and consider their shadow (orthogonal projection) on a plane,  $\Pi$ , parallel to the cylindrical axis. The result is a family of curves, two through each point, all lying between the two parallel lines  $L_1, L_2$ , which bound the projection of the cylinder ( $L_1$  and  $L_2$  are the intersections with  $\Pi$  of the two planes perpendicular to  $\Pi$  and tangent to the cylinder). All the projections on  $\Pi$  of the helices (actually, sine curves) are tangent to  $L_1$  and  $L_2$ : these lines are their envelopes and bear the same relation to them as caustics do to rays. This shows that a set of curves (the helices) which have no "peculiarity" at all on a surface (the cylinder) upon which they lie, may be represented in projection by curves (the sine curves between  $L_1$  and  $L_2$ ) which do have marked peculiarities.

The fundamental question is whether such peculiarities, as they occur in caustics of acoustic rays, have any significance in the

physics of propagation of acoustic energy--which is, of course, the only reason for using rays in the first place. An essential point of the present study is to show that if the rays are considered in phase space, they can be used directly to calculate the propagation of energy without the difficulties of caustics, etc., which difficulties are reduced to "artifacts of the representation." But this requires certain concepts which must now be set forth.

It is important to realize that our methods skip the wave-front and Huygens construction of Sections 1 and 2, as well as the eikonal--just as modern dynamics uses Hamiltonian theory but not the Hamilton-Jacobi partial differential equation (except locally or in simple cases). Its use will be extended and clarified in Section 16.

We turn lastly to the special case of Section 7, when  $c = c(x, z)$ , and write the equations of motion of a particle constrained to lie on the smooth surface  $\Sigma$  constructed as described there. If  $X, Y, Z$  are the components of the forces of constraint, the equations of motion are

$$\frac{d^2x}{d\tau^2} = X, \quad \frac{d^2y}{d\tau^2} = Y + g, \quad \frac{d^2z}{d\tau^2} = Z;$$

and, of course, since the particle must lie on the surface  $\Sigma$ , we must have (introducing the abbreviation  $f(x, z)$ ):

$$y = f(x, z) = -V/g = C^2/2c^2g$$

Finally, the forces of (smooth) constraint result from the normal reaction of  $\Sigma$ , in an upward direction; i.e., opposite to that of the

+ y - axis. Since the direction components of this normal are  $f_x$ , -1,  $f_z$  (first partials of  $f(x, z)$ ) we have

$$X = Rf_x, \quad Y = -R, \quad Z = Rf_z,$$

where  $R$  is the force of reaction and is positive. Substituting the values given by the last two sets of equations into the first, we obtain

$$\frac{d^2x}{d\tau^2} = Rf_x, \quad \frac{d^2}{d\tau^2} f(x, z) = -R + g, \quad \frac{d^2z}{d\tau^2} = Rf_z.$$

Carrying out the differentiations in the second of these equations we have

$$\frac{d}{d\tau} f(x, z) = f_x \frac{dx}{d\tau} + f_z \frac{dz}{d\tau}$$

$$\frac{d^2}{d\tau^2} f(x, z) = f_x \frac{d^2x}{d\tau^2} + f_z \frac{d^2z}{d\tau^2} + f_{xx} \left( \frac{dx}{d\tau} \right)^2 + 2f_{xz} \frac{dx}{d\tau} \frac{dz}{d\tau} + f_{zz} \left( \frac{dz}{d\tau} \right)^2$$

$$= R(f_x^2 + f_z^2) + \quad + \quad + \quad + \quad +$$

This shows that the vertical component of the acceleration on  $\Sigma$ , i.e.,  $d^2y/d\tau^2$ , being a linear combination of the second derivatives and the squares of the first derivatives of  $f$  (which is proportional to  $1/c^2$ ), can be neglected. Hence, to this approximation,  $R = g$ , which converts the remaining equations of motion into

$$\frac{d^2x}{d\tau^2} = gf_x = -\frac{\partial V}{\partial x}, \quad \frac{d^2z}{d\tau^2} = gf_z = -\frac{\partial V}{\partial z},$$

As proved before, these are the ray equations in the dynamical form (8.2).

## 9. Differential Systems and Their Integral Invariants\*

A recurring feature in our study of lines of power flow, rays and trajectories in phase space, is the system of differential equations of the first order such as (3.5) (and the previous one, containing  $dt$ ), (5.1), and (8.4). These all come under a general form, which can be written symmetrically as

$$(9.1) \quad \frac{du_1}{U_1} = \frac{du_2}{U_2} = \dots = \frac{du_n}{U_n}$$

where  $U_1, \dots, U_n$  are given functions of the  $n$  variables  $u_1, \dots, u_n$ . If it is wished to identify the independent variable among these  $n$  variables, we may re-write (9.1) in a less symmetrical form. If, for example,  $U_n \neq 0$ , we may use  $u = u_n$  as the independent variable, and set  $V_i = U_i/U_n$ . Then (9.1) becomes

$$(9.2) \quad \frac{du_i}{du} = V_i = V_i(u_1, \dots, u_{n-1}, u) \quad (i=1, \dots, n-1).$$

In the form (9.1) the system has a simple geometrical interpretation, easily visualized when  $n = 2$  (the plane) or  $n = 3$  (ordinary space). The denominators (assumed not all simultaneously zero) can be thought of as determining the direction of a line segment at each point  $(u_1, \dots, u_n)$ , namely, the segment whose axial projections (direction numbers) are in the proportions  $U_1 : U_2 : \dots : U_n$ . The solutions of (9.1), or "integral curves", are any curves in the space of the  $n$  variables, tangent at every one of their points to the line segment given there. In other words, (9.1) determines a field of directions; an integral curve is one having these directions as those of its tangents. The fundamental property of these

\*cf., e.g., Goursat & Hedrick, Vol. II.<sup>4</sup>

curves is that, through any given point, passes one and only one integral curve: they decompose space into "filaments". This is a theorem which presupposes a "regularity" of the denominators of (9.1) in the region considered: the continuous differentiability of the sort that is always safely assumed in physical applications. It also assumes that at no point of our region are they all equal to zero.

In the form (9.2) the system has a simple kinematic interpretation, in terms of motion in the  $(n-1)$ -dimensional space of the dependent variables  $(u_1, \dots, u_{n-1})$ . If we interpret the independent variable  $u$  as "time", the derivatives in (9.2) are the velocities of a moving point in this space, so that the system (9.2) assigns a velocity vector of components  $V_1, \dots, V_{n-1}$  to each point  $(u_1, \dots, u_{n-1})$  and time  $u$ . The fundamental property stated in the preceding paragraph now takes the following form: Given any point  $(u_1^0, \dots, u_{n-1}^0)$  and time  $u^0$ ; then there exists one and only one path, or trajectory, described by a moving point  $(u_1, \dots, u_{n-1})$ , whose coordinates are functions of  $u$  satisfying (9.2), and which, when  $u = u^0$ , is in the position given by  $(u_1^0, \dots, u_{n-1}^0)$ . In formulas, this means that there are  $n-1$  functions  $f_i$  of the  $n$  given constants  $u_1^0, \dots, u_{n-1}^0, u^0$ , and also  $u$ , such that, on setting

$$(9.3) \quad u_i = f_i(u_1^0, \dots, u_{n-1}^0, u^0, u) \quad (i=1, \dots, n-1)$$

and interpreting  $du_i/du$  as the partial derivative of  $f_i$  with respect to  $u$  (the  $0$ -superscript variables held constant), the equations (9.2) are true for all values of all  $n+1$  variables (restricted to lie in the region of regularity of the functions  $V_i$ ).

In the case  $n=4$  our kinematic interpretation of (9.2) may be considered as picturing the flow of a fluid, which fills the region of

space of  $(u_1, u_2, u_3)$ , and moves so that, at each moment  $u$ , the field of velocities of its particles is given by (9.2). Then (9.3) can be regarded as the transformation experienced by any point  $P$  or figure (locus of points) traced in the fluid (marked, e.g., in ink or smoke) and partaking of its motion, as time  $u$  goes on. While not visible physically when  $n > 4$ , it is convenient to apply the terminology of "flow" in all cases. In fact, all the essential properties to be used later that are intuitively evident for  $n \leq 4$  can be established mathematically for the general case.

Analytically, the effect of the flow transformation induced by the solution (9.3) of (9.2) upon figures traced in the fluid can be found by the following general method. If the figure is a curve whose initial position  $C_0$  is given parametrically by setting the  $n-1$  coordinates equal to functions of a parameter  $\alpha$ , during the course of the flow its successive images are also given parametrically in terms of  $\alpha$  by (9.3) after each  $u_1^0$  has been replaced by the corresponding function of  $\alpha$  initially given. The points of  $C$  will then be functions of two variables: time  $u$  and position along  $C$ :  $\alpha$  (of course  $u^0$  is held fast). Similarly, if a surface  $S$  is given parametrically in its initial position with the aid of two parameters  $\alpha, \beta$ , to find its parametric equations in its evolving positions, one simply replaces  $u_1^0$  in (9.3) by the given function of  $\alpha$  and  $\beta$ . For a region  $V$  of dimension  $n-1$  (a piece of ordinary space when  $n=4$ ) one applies (9.3) with  $(u_1^0, \dots, u_{n-1}^0)$  ranging over the given initial position of this region,  $V_0$ .

A first integral of (9.1) is a (regular) function  $F=F(u_1, \dots, u_n)$  which is constant along every integral curve. This is the same as saying

that its differential  $dF$  is zero along such a curve; or, applying (9.1), that  $F$  satisfies the partial differential equation

$$(9.4) \quad U_1 \frac{\partial F}{\partial u_1} + \dots + U_n \frac{\partial F}{\partial u_n} = 0.$$

In geometric language, this means that the gradient  $\nabla F$  (of components  $\partial F / \partial u_i$ ) is perpendicular to the direction  $U_1 : U_2 : \dots : U_n$ . Equivalently, it means that the locus  $F = K$  (constant) is tangent to that direction. Also, from the defining property of  $F$ , any integral curve of (9.1) having one point on the locus  $F = K$  lies entirely in this locus: we can then say that the locus is "made up" of integral curves. Conversely, if  $n-1$  first integrals  $F_i$  produce loci  $F_i = K_i$  which intersect in a curve, this curve is evidently an integral curve. It is shown mathematically that every integral curve in our region can be given as such an intersection.

The first integral  $F$ , in the notation of (9.2) and its interpretation as a flow in  $(n-1)$ -dimensional space, can be described as a function of position and time whose rate of change along any trajectory is zero. This is because, in the notation in question, (9.4) may be written as

$$(9.5) \quad V_1 \frac{\partial F}{\partial u_1} + \dots + V_{n-1} \frac{\partial F}{\partial u_{n-1}} + \frac{\partial F}{\partial u} = 0;$$

and the left-hand side of this is precisely the derivative of  $F$  along the trajectory, since by the formulas of partial differentiation

$$(9.6) \quad \frac{dF}{du} = \frac{\partial F}{\partial u_1} \frac{du_1}{du} + \dots + \frac{\partial F}{\partial u_{n-1}} \frac{du_{n-1}}{du} + \frac{\partial F}{\partial u},$$



the values of  $du_1/du$  being given by (9.2). This has the following interpretation (phrased for the case  $n=4$ , but easily imagined in general). If initially ( $u=u^0$ ) the region of fluid for which  $F \leq K$  is colored by a dye, this colored region will change in shape and position during the flow, i.e., as the time  $u$  increases; but it will always be bounded by the surface  $F=K$ : it cannot cross this boundary. This surface will move and deform in the course of the flow, except when  $F$  does not contain  $u$ : then it is a fixed boundary.

In the important special case that the functions  $V_i$  are independent of time  $u$ , the flow is stationary: the field of velocities in the  $(U_1, \dots, U_{n-1})$  space is unchanging, and the trajectories are fixed curves.

If  $V$  is an  $(n-1)$ -dimensional portion of the medium, given as  $V_0$  when  $u=u^0$ , what becomes of its volume (number of cubic units when  $n=4$ , and generalized) as  $V$  evolves in the course of the flow? If the flow is incompressible, the volume of  $V$  will not change. If the flow is conservative, the mass will not change. This means that there is a density function,  $\rho = \rho(u_1, \dots, u_{n-1}, u)$ , which has the property that for all choices of initial volume  $V_0$

$$(9.7) \quad \int_V \rho \, du_1 \dots du_{n-1} = \int_{V_0} \rho_0 \, du_1^0 \dots du_{n-1}^0,$$

where  $\rho_0$  denotes the result of replacing  $(u_i, u)$  by  $(u_i^0, u^0)$ . By a theorem of the calculus (explained in more detail in Appendix E), for this to occur,  $\rho$  must satisfy the "equation of continuity"

$$(9.8) \quad \frac{\partial}{\partial u_1} (\rho V_1) + \dots + \frac{\partial}{\partial u_{n-1}} (\rho V_{n-1}) + \frac{\partial \rho}{\partial u} = 0$$

This reduces to

$$(9.8') \quad \frac{\partial V_1}{\partial u_1} + \dots + \frac{\partial V_{n-1}}{\partial u_{n-1}} = 0$$

in the special case of incompressible flow, because then  $\rho$  is a constant and drops out.

We have already seen examples of those equations of continuity (with  $n=4$ ) in Sections 3 and 5, where it was not matter but energy that was conserved. We now apply the ideas to the ray differential equations in their Hamiltonian form (8.4). In this case  $n=7$  and the line-up of variables is the following

$$\begin{array}{ccccccc} x & y & z & p_x & p_y & p_z & \tau \\ u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u \end{array}$$

while that of the given functions is

$$\begin{array}{cccccc} \partial H / \partial p_x & \partial H / \partial p_y & \partial H / \partial p_z & - \partial H / \partial x & - \partial H / \partial y & - \partial H / \partial z \\ V_1 & V_2 & V_3 & V_4 & V_5 & V_6 \end{array}$$

On substituting these values into (9.8') it becomes evident that the equation is satisfied: the flow that (8.4) induces in phase space is incompressible. This is Liouville's theorem, valid for every Hamiltonian system, and forming an essential basis of statistical mechanics.

We have seen in Section 8 that the function  $H$  is a first integral of (8.4), and that the acoustic rays form the locus  $\mathcal{S}: H=0$ . This locus is a fixed one in phase space, since  $H$  is independent of  $\tau$ . The trajectories in phase space lying in this locus induce a flow in it, but

taking place in a space of 5 dimensions rather than in 6. Now whereas the flow in this space is not in general incompressible (5-dimensional volumes change), there is a positive density function  $\rho$ , depending only on position on  $\mathcal{S}$ , and  $\rho$  satisfies the condition for conservation (9.7), but with  $V$  and  $V_0$  5-dimensional regions of  $\mathcal{S}$  and  $du_1 \dots du_{n-1}$ , etc., replaced by the "element of 5-volume" for this locus. The proof of this and other related theorems will be taken up in Appendix E, using powerful general methods. We limit ourselves here to giving an intuitive argument applying to a parallel case in ordinary 3-dimensional space in which a steady flow of an incompressible fluid (liquid) is taking place, leaving  $F$ , a function of position, invariant (i.e.,  $F$  is a first integral of the equations defining the steady flow).

Let  $\mathcal{S}_2$  be the surface  $F=0$ , and consider the near-by surface  $\mathcal{S}_2' : F=h$  (for any small constant  $h$ ). These are fixed surfaces, very close together, as shown in Figure 9.I. Think of a piece  $\Delta \mathcal{S}_2$  of  $\mathcal{S}_2$  and draw the normals at its boundary, cutting off a piece  $\Delta \mathcal{S}_2'$  of  $\mathcal{S}_2'$ , all forming a cylinder-like region  $\Delta V$ . During the flow,  $\Delta V$  will move,

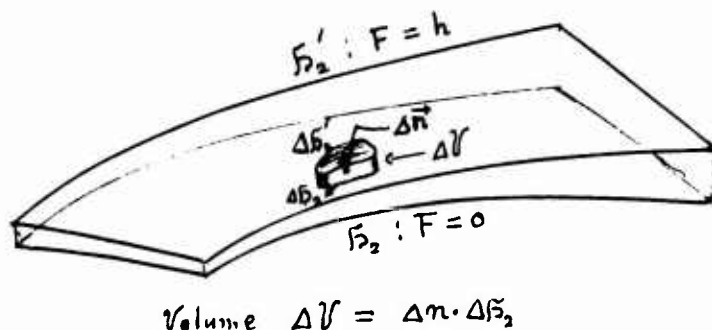


FIGURE 9.I

$\Delta \mathfrak{H}_2$  and  $\Delta \mathfrak{H}_2'$  sliding along  $\mathfrak{H}_2$  and  $\mathfrak{H}_2'$ , but its volume will remain constant. It will be bounded by the moving pieces  $\Delta \mathfrak{H}_2$ ,  $\Delta \mathfrak{H}_2'$ , and their connecting cylinder-like surface. For small dimensions of the figure and a short interval of time, we can regard  $h/\Delta n$  as the magnitude of the gradient of  $F$ , so that  $\Delta n = h/|\nabla F|$ , to terms of the first order. To the same degree of approximation, the volume of  $\Delta V = \Delta n \cdot \text{area of } \Delta \mathfrak{H}_2 = \text{area of } \Delta \mathfrak{H}_2 / |\nabla F|$ ; i.e., if the numerator is regarded as an element of surface  $dS = \text{area of } \Delta \mathfrak{H}_2$ , we see that the element of surface integration,  $dS/|\nabla F|$ , is unchanged by the flow, so that

$$\iint_S \rho \, dS = \iint_{S_0} \rho \, dS_0, \quad \rho = 1/|\nabla F|.$$

In the application to the Hamiltonian case, a corresponding equation holds, but with

$$\rho = 1/|\nabla H| = \left[ H_x^2 + H_y^2 + H_z^2 + H_{p_x}^2 + H_{p_y}^2 + H_{p_z}^2 \right]^{-1/2}.$$

Applying the equations of Section 8, this reduces to

$$\rho = 1/n (1 + |\nabla n|^2)^{1/2}, \quad (n = 1/c).$$

Quantities that are left unchanged by certain transformations (as by the flow (9.3) induced by (9.2)) are called invariants of these transformations. A function such as a first integral  $F$  is an invariant function of position  $(u_1, \dots, u_{n-1})$ ; the mass of a piece of matter undergoing a conservative flow is also an invariant quantity, but is not a function of position but an integral over a figure which is moved and deformed by the flow: such a quantity is called an integral invariant of

the system of equations (or of the transformations they induce). The integral of  $E$  in Section 3 that expresses the acoustic energy in a volume of the medium is an integral invariant of the system of power flow lines. The volume of an incompressible fluid, the "6D" volume of Hamiltonian phase space and the "5D" surface integral of  $1/|\nabla H|$  in the 5-dimensional ray phase space  $\mathfrak{H}$  are all integral invariants.

There is a 4-dimensional integral invariant of importance to the present study of acoustic propagation; it may be obtained from the one mentioned at the close of the last paragraph. The intuitive explanation is very simple, and will be given in the analogous case of steady flow in ordinary space, deferring mathematical details to Appendix E.

Figure 9.II shows a set of trajectories--fixed stream-lines, since the flow is stationary--forming a tube-like surface,  $\mathcal{G}$ , closed by two surfaces,  $A$  and  $B$ .  $A$  is any smooth surface closing the tube, while  $B$  is the result of sliding  $A$  continuously along the stream-lines. This means that each point  $P$  of  $A$  is moved through a distance  $D(P)$ , this

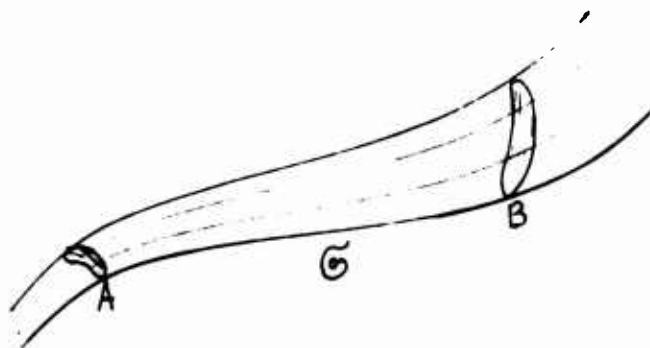


FIGURE 9. II

distance being a regular function of position. Clearly, the net mass (measured by the density  $1/|\nabla F|$ ) which enters this two-ended tube is zero because of the steady nature of the flow. Also, nothing can pass through the tubular boundary; therefore what enters the tube through A must equal what leaves it through B. But the amount which crosses any surface fixed in the fluid is the surface integral of the normal component of the flux vector (density times velocity vector, of components  $\rho u, \rho v, \rho w$ ). Therefore this integral taken over A is equal to the same one over B. If B is chosen as a surface into which A flows during a given interval of time, the fact that the flux integral is a 2-dimensional integral invariant is obvious; but B may be obtained by sliding A in other ways.

But our example gives more than an invariant integral over a surface that moves with a flow: it gives us an integral that has a much higher degree of invariance--that may be described as a "sliding integral invariant."

While more details will be given in Appendix E, we may add to our examples one of a rather different character, the circulation along a curve C in the physical space in which a flow of a fluid is taking place. If the components of the hydrodynamic velocity are  $u, v, w$ , the circulation along it, in a given direction, is defined as the line integral of the tangential component, in that direction, of the velocity:

$$\text{circulation} = \int_C u \delta x + v \delta y + w \delta z$$

Here  $\delta$  is used to distinguish differentiation along the curve (time held fixed) from  $d$  used for differentiation along the trajectory (curve parameter  $\alpha$  held fixed). A fundamental theorem in hydrodynamics states that for a "perfect fluid" the circulation about every closed curve C is an

integral invariant of the flow. Upon this theorem is based Lagrange's theorem of the permanence of the irrotational character of a perfect fluid initially irrotational, and thence the existence of a velocity potential. And again, the Helmholtz theory of vortices, the "vorticity" vector field resulting from the circulation through Stokes' theorem.

These circumstances have their parallel in Hamiltonian Theory, where the "action integral"

$$\int_C p_x \delta x + p_y \delta y + p_z \delta z$$

is an integral invariant--provided  $C$  is closed. As will be explained in Appendix E, from this all the other integral invariants of the Hamiltonian system can be derived by simple formulas.

In Appendix E this whole subject of integral invariants is studied in mathematical detail, and references to the literature are given.

## 10. Energy Flow in Phase Space

Having, as explained in Section 8, "lifted" the rays of the physical space of the medium into the 5-dimensional phase-space  $\mathcal{S}$ , and having shown in Section 6 that in a Fermat family of travelling waves  $\{\phi_\omega\}$  have their power flow or  $\Lambda$ -lines approaching coincidence (as  $\omega \rightarrow \infty$ ) with an associated class of rays or L-lines, it is necessary to make clear the implications of these facts for the phase space picture--both as regards its geometry and its relation to the flow of power.

The first step is to "lift" each  $\Lambda$ -line of our Fermat family into the 5-dimensional  $\mathcal{S}$  as follows: For each value of  $\omega$  and point  $P: (x, y, z)$  on a  $\Lambda$ -line (not on  $\Gamma$ ) we have a unique direction  $\vec{U}_\omega$  tangent to it, in the sense of the flow of power, as explained in Section 6. If  $(\cos \alpha, \cos \beta, \cos \gamma)$  are the components of this vector (its direction cosines), we define the three momenta by the equations  $p_x = n(x, y, z) \cos \alpha$ ,  $p_y = n(x, y, z) \cos \beta$ ,  $p_z = n(x, y, z) \cos \gamma$ ; or, in vector notation, by  $\vec{p} = n \vec{U}_\omega$ . Clearly, the point  $(x, y, z, p_x, p_y, p_z)$  in 6-dimensional phase-space lies in the 5-dimensional  $\mathcal{S}$ , since its coordinates as just defined satisfy  $H = 0$ . Evidently its orthogonal projection onto the space of the medium is the point  $P: (x, y, z)$  from which we started. By applying this process to all points of a  $\Lambda$ -line, it is lifted into a curve in  $\mathcal{S}$ . In view of the smoothness of the original  $\Lambda$ -line, its lift also is smooth. Of course it will not in general be an integral curve of the canonical equations of the rays (8.4); but as  $\omega \rightarrow \infty$  it will approach coincidence with its associated tangent ray at  $P$ .

Now let  $\mathcal{G}$  be a small piece of surface in the medium, nowhere tangent to a limiting ray  $L$  of the Fermat family--and hence, for sufficiently



great  $\omega$ , nowhere tangent to any of its  $\Lambda$ -lines. We wish to study the flow of power across  $\mathcal{G}$  by the use of rays---a method easy to explain for short ranges of propagation in terms of the constructions of Section 1, together with the approach of  $\Lambda$ - to L-lines, but breaking down when caustics, etc., appear. Our first step is to lift  $\mathcal{G}$  into  $\mathcal{H}$  by first defining a field of direction vectors, just one at each point of  $\mathcal{G}$ , and then assigning as momentum vector the product of its direction cosines by the values of  $n$  at the respective points of  $\mathcal{G}$ . Of course, this can be done in as many different ways as there are ways of defining these fields of directions: the resulting lifts will be 2-dimensional figures in the 5-dimensional phase space, two of which can only intersect at points at which the corresponding directions in their fields happen to coincide.

An important case of this construction occurs when the field of directions assigned to points of  $\mathcal{G}$  are those of the limiting rays  $L$  through these points: we would obtain a perfectly definite lift, which we shall denote by  $\mathcal{G}^L$ . If on the other hand we use the directions of the  $\Lambda$ -lines cutting  $\mathcal{G}$  we obtain (for each  $\omega$ ) another non-intersecting lift  $\mathcal{G}^\Lambda$ . The same will be true of the 3-dimensional figure obtained by lifting all the limiting rays that cut  $\mathcal{G}$ ---and hence whose lifts cut  $\mathcal{G}^L$ ---and also the 3-dimensional lift of all the  $\Lambda$ -lines (for the given  $\omega$ ) which cut  $\mathcal{G}$  and therefore whose lifts cut  $\mathcal{G}^\Lambda$ . Any intersections of these two 3-dimensional "worm-like" figures in the 5-dimensional  $\mathcal{H}$  will be "purely fortuitous"---as will intersections of two  $\Lambda$ -line "worms" for different values of  $\omega$ . These statements may have exceptions in special cases, as in the third and fourth examples of Appendix D where  $\mathcal{G}^L$  and  $\mathcal{G}^\Lambda$  coincide.

We may actually visualize the geometrical relations in the case of waves in the vertical  $xz$ -plane, to be studied at length in Sections 13-16. Then the phase-space  $\mathcal{G}_3$  is 3-dimensional (shown in Figures 14-I, II);  $\mathcal{G}$  is replaced by a piece of curve, and the lifts  $\mathcal{G}^L$  of the rays cutting it, and also the  $\mathcal{G}^\Lambda$ , are ribbon-like figures, in general never intersecting.

In spite of the fact that we do know that the worm of lifted  $\Lambda$ -lines cutting  $\mathcal{G}^\Lambda$  approaches that of the  $L$ -lines through  $\mathcal{G}^L$  as  $\omega \rightarrow \infty$ , this in itself gives but an awkward handle on the power flux relations in  $\mathcal{H}$ . To obtain a better hold on the question, we shall make the following modification of the construction--simple but drastic in its effects:

What we shall do is, simply, to increase by two units the dimensionality of  $\mathcal{G}^L$ : at each point  $P$  of  $\mathcal{G}$  we take not only the unique direction of  $L$  there, but a cone of solid angle  $\Omega$  of directions about it, and, after multiplying each set of direction cosines by the value of  $n$  at  $P$ , obtain a set  $\Pi$  of momenta lying in this cone. Obviously this  $\Pi$  is a 2-dimensional piece, and when combined with each point of the 2-dimensional  $\mathcal{G}$ , gives a 4-dimensional lift in  $\mathcal{H}$ .

A simple and convenient way of making this construction is first to pick a reference point  $P_0$  inside  $\mathcal{G}$ ; second, to fix a cone  $\Omega_0$  containing the direction  $\vec{U}_0$  of the  $L$ -line through  $P_0$  in its interior; and third, to take for the  $\Omega$  at any other point  $P$  a cone congruent and parallel to  $\Omega_0$ : if the dimensions of  $\Omega$  are small enough in comparison with these of  $\Omega_0$ , as we will suppose, for each  $P$  on  $\mathcal{G}$ , the  $\Omega$  will contain the  $\vec{U}$  at  $P$  in its interior. We may always represent any such cone

of directions as a region on the unit sphere centered at its vertex  $P$ ; its spherical area will be the solid angle measured in steradians. Such a combination of objects  $(P, \vec{U})$  is an element of the Cartesian product  $\mathcal{G} \times \Omega$ .

The corresponding combination  $(P, n(P)\vec{U})$  is a point in  $\mathcal{H}$ , and their totality is the 4-dimensional lift of  $\mathcal{G}$  that we are defining. We shall denote it by  $(\mathcal{G}, \Pi)$ ; only when  $n$  is constant throughout  $\mathcal{G}$  is it a Cartesian product  $(\mathcal{G}, \Pi) = \mathcal{G} \times \Pi$ . Nevertheless, when both  $\mathcal{G}$  and  $\Omega$  are very small, the measure (4-dimensional volume) is, to quantities of higher order, given by the following formula, where  $n_0$  is the value of  $n$  at  $P_0$ :

$$(10.1) \quad \text{measure } (\mathcal{G}, \Pi) = n_0^2 \text{ measure } (\mathcal{G} \times \Omega) = n_0^2 \text{ area of } \mathcal{G} \times \text{area of } \Omega$$

This formula will be used in Section 11.

Now consider the rays determined by all the points of  $(\mathcal{G}, \Pi)$ : they are the integral curves of (8.4) passing through these points, and, when extended in range, form an elongated 5-dimensional piece of  $\mathcal{H}$ , to be denoted by  $\mathcal{H}(\mathcal{G}, \Pi)$ . Since we are thinking of both  $\mathcal{G}$  and  $\Omega$ --and therefore  $\Pi$ --as small in largest dimension, this piece is not only long but slender: it may be described as a "hyper-worm".

Since every direction of the  $L$ -lines at points of  $\mathcal{G}$  are interior to the corresponding cones  $\Omega$ , it follows that the 2-dimensional lift  $\mathcal{G}^L$  is interior to the 4-dimensional  $(\mathcal{G}, \Pi)$ , and hence the 3-dimensional "worm" formed by the lifted  $L$ -lines through it is inside the above 5-dimensional hyper-worm.

Suppose that the range interval through which the rays (i.e., their lifts) have been extended is thought of as large, but fixed. Since as  $\omega \rightarrow \infty$  the worm of  $\Lambda$ -lines through  $\mathcal{G}^\Lambda$  approaches that of the L-lines through  $\mathcal{G}^\Lambda$ , and since they form a set of interior points to  $\mathcal{H}(\mathcal{G}, \Pi)$ , (strictly interior, if the rays in the construction of the latter are extended very slightly at each end), it follows that for all sufficiently great  $\omega$  the whole worm of  $\Lambda$ -lines through  $\mathcal{G}^\Lambda$  will be inside  $\mathcal{H}(\mathcal{G}, \Pi)$ .

This attaches a power flow number to the hyper-worm  $\mathcal{H}(\mathcal{G}, \Pi)$ : the acoustic power entering the  $\Lambda$ -worm through  $\mathcal{G}$ , once the frequency has reached its value as required above. This power can be thought of as flowing along the hyper-worm, and crossing any of its "cross-cuts": the 4-dimensional figures in  $\mathcal{H}(\mathcal{G}, \Pi)$  which divide it in two pieces in such a way that each ray forming it is cut at an angle into two pieces; and, further, that their (topological) position in the hyper-worm is such that they can be slid continuously along the rays that make it up into the original figure  $(\mathcal{G}, \Pi)$ . In Appendix G an analytic basis will be given for the statements here submitted on an intuitive one (which, in general cases must be replaced by rigorous analysis of limits). Let the projection of the cross-cut into the space of the medium, in which  $(x, y, z, p_x, p_y, p_z)$  becomes  $(x, y, z)$  with the direction determined by  $(p_x, p_y, p_z)$  through it, is "in general" a solid (shaped, e.g., like a potato), through each point of which passes a continuous 1-parameter family of directions, describable as a narrow curved "fan". The limiting L-lines traverse this potato, being at each point tangent to the corresponding fan. Second, it is possible to slide this cross-cut along the

rays that cut it so that it assumes a special position, exactly resembling the original  $(\mathcal{G}, \Pi)$ : The fans moving together along their rays have squeezed into cones of small solid angle and the potato has become a surface  $\mathcal{G}'$  like  $\mathcal{G}$ . All the power entering the hyper-worm through  $\mathcal{G}$  crosses  $\mathcal{G}'$ : this is our power flow number--obviously a sliding invariant.

This is the first step in the transference to phase space of the power flux quantities. It gives us a set function, i.e., a numerically valued function of 4-dimensional pieces of  $\mathcal{S}$  that can be regarded as cross-cuts; and since any small enough piece in this space which is nowhere tangent to a ray can be regarded as a cross-cut (i.e., of the hyper-worm made of all the rays that cut it), this includes a large class of figures. We have seen that this set function is a sliding invariant of the rays in phase space. It is easy to see that it is "additive": the flow of power through both of two cross-cuts, which cannot be slid along their rays so as to overlap, is the sum of the flows through each.

But this sliding invariant additive set function is not an invariant integral--i.e., of a density over the set in question. This is because it is not "absolutely continuous"--the power flow across  $(\mathcal{G}, \Pi)$  could remain constant and positive while the measure of this set approaches zero. To see this we have but to recall (10.1), and let  $\Omega$  shrink up about each  $L$ -line through  $\Omega$ , the latter remaining constant: such "density" over the 4-dimensional cross-cut would jump from 0 to  $\infty$ .

We are to regard this invariant set function as a "precursor" of an actual sliding integral invariant; but this requires the introduction of infinite classes of Fermat families, acting together but with enough random to avoid the rigidity of the "principle of superposition". This will form the subject of the next section.

## 11. The Statistics of Acoustic Transmission

In all studies of radiant energy two types of model may be used: firstly, the single solution  $\psi$  of the wave equation under quite specific boundary and initial conditions; and secondly, an aggregate or ensemble of such solutions, combining "incoherently", and producing the observed effects and regularities through the statistical characteristics of the ensemble, as well as the properties of its individual members. In electromagnetic radiation, the first type of model is appropriately applied to radio and radar waves, and recently to laser radiation under simple conditions; the second type of model becomes necessary in most studies of light, particularly when the source is large in comparison with the wavelength, and in the isotropic radiation in cavities. In acoustics, the first type of model is appropriate in studies in which the configuration of the environment is simple and its dimensions not inordinately greater than the wavelengths considered, as in room acoustics, lobe formation and other interference patterns near reflecting objects, and in many similar cases. On the other hand, the second type of model--the statistical ensemble of waves--is necessary to describe the actualities of hydroacoustic radiation of wavelengths much shorter than the ranges, the depth of the medium, and the radius of curvature of the rays and power flow lines. This choice of model is not just a matter of convenience--e.g., to obtain a tractable set of relations--but is forced upon us by those aspects of the physical situation now to be outlined, which subject them to random.

The first source of random, contributing a set of incoherent sound waves, is the nature of the acoustic emitters of interest to Naval

operations, such as submarines. Such an object is large and has many vibrating parts. Even the pressure hull has a complicated pattern of vibration, distant parts seldom vibrating in perfect unison. Added to this is the irregular nature of the mass of water through which it passes and which takes part in the acoustic emission, e.g. by cavitation and turbulence.

The second source of random is the medium--the actual behavior of the sea water. There is insufficient knowledge of its effects: while a single acoustic wave has been shown to maintain a remarkable degree of phase coherence over long ranges, there is no indication that two waves, generated by different parts of the emitter, will maintain any observable degree of mutual coherence. cf. Tatarski<sup>2</sup> and others' discussion of random turbulence.

The third source of random in the acoustic radiation fields is the set of physical boundaries. The simplest is the ocean surface. While there is much evidence to regard it as acting as a "specular reflector" (in the range of frequencies and angles of interest), this term implies two properties: the return of all acoustic energy without appreciable loss at an angle of reflection equal to the angle of incidence; and the maintenance of a definite phase relation between incident and reflected wave, so that energies combine by interference rather than simple addition. It is only in the first sense that the ocean surface appears to act as a specular reflector. A more complicated boundary is the bottom, which absorbs a fraction of the incident radiation and reflects the remainder. With hard bottoms and moderate ranges, the latter part is important as contributing to detection by "bottom bounce"; but it has a randomizing effect on what it returns. In fact, recent studies have been made of the influence of bottom irregularities on the random nature of the returned radiation.<sup>2</sup> In the case of

very long range propagation, the repeated bottom absorptions of certain rays remove all their energy contribution, so that the bottom assumes the character of a perfect absorber. This, however, implies an extreme regularity of the rays, the medium, and the bottom: it may be necessary to take into account rays that are reflected only once by the bottom, (e.g., at a sea-mount) and then enter the radiation field in a phase and direction uncorrelated with the others.

There is a final reason for the appropriateness of statistical methods in the treatment of hydro-acoustic propagation: the limitations of possible specific knowledge. The fine-structure of the environment--sound speed, all physical boundaries, etc.--have excessively complicated and essentially unpredictable variations in space and time. All that can ever be known of these features is in terms of averages, too coarse-grained to form a basis of more than a statistical treatment. Fortunately, when properly applied, this gives all that is needed for practical operations.

On the basis of these considerations, we have to formulate in precise terms the statistical assumptions regarding the radiation field to be used as the model of hydro-acoustic transmission. The ideas are essentially simple--given the material set forth in the preceding sections.

Suppose that  $\Psi = \{\psi_1, \dots, \psi_N\}$  is a set of elementary travelling waves of common frequency  $\omega$ , all satisfying the loose boundary conditions appropriate to the practical problem. Each  $\psi_i$  has its own vector field  $\vec{F}_i$  and corresponding power flow lines. We construct the small pieces  $\Delta G$ ,  $\Delta \Omega$ , of surface  $G$  and of solid angle  $\Omega$  (nowhere tangent to this surface), as in the preceding section. The power lines of  $\psi_i$  may or may not have a bundle crossing  $\Delta G$  in a direction belonging to  $\Delta \Omega$ . If they do,  $\psi_i$  will contribute



a certain quantity of acoustic power to the "hyper-worm"  $K = \mathfrak{S}(\Delta G, \Delta \Omega)$ : once entering it, it will stay in it, as explained before.\* The first part of our statistical assumption is that all these waves are incoherent in phase; hence their total power contributions add. Let  $P(\Psi)$  be the algebraic sum of all the power contributed by the set  $\Psi$  to this  $K$  (algebraic sum, since the directions involved in the above construction may show negative power entering--i.e., positive power leaving, our  $K$ ). For the reasons given before, this total power entering  $K$  through its base figure  $(\Delta G, \Delta \Omega)$  stays in it, and is what traverses any later cross-cut (what we get out of the worm is just what we feed it): it is a "sliding invariant"; but as yet, not expressible as an integral. This has all been explained in Section 10.

The next step in the formulation removes the discreteness and resulting discontinuities in power flow directions due to confining the ensemble  $\Psi$  to a finite ~~sum~~. The method is to use limits of such sums of elementary travelling waves, as their number increases indefinitely while their individual energies approach zero. The ideas are comparable to those in the definition of an integral as a limit of a sum. To put the matter precisely, we think of an infinite "triangular" collection of elementary travelling waves (of the same frequency  $\omega$ ). Each one is identified by two indices, as  $\psi_i^N$ , in which  $i$  goes from 1 to  $N$ , giving a "horizontal" set or row of  $N$  waves; while  $N$  increases indefinitely, giving an infinitely deep "pile" of such horizontal rows, and thus the following triangular figure:

---

\* This actually assumes that each  $\psi_i^N$  is in reality a Fermat family  $\{\psi_i^N(\omega)\}$  and that  $\omega$  is sufficiently great so that all the  $\Lambda$ -lines are in  $K$ . Increasing  $\omega$  along the perpendicular to the page, our triangle  $\Psi$  becomes an infinitely tall prism, with an infinite triangular base--from which we then cut off that part with too low values of  $\omega$ .

$$\begin{array}{ccc}
 \psi_1^1 & & \psi^1 \\
 \psi_1^2, \psi_2^2 & & \psi^2 \\
 \Psi: & \dots & . \\
 \psi_1^N, \psi_2^N, \dots, \psi_N^N & & \psi^N \\
 & \dots & .
 \end{array}$$

Corresponding with this triangular system of waves, we have a triangular system of their energy densities  $E_i^N$  and of their power flux vectors  $\vec{F}_i^N$ , where, as with  $\psi_i^N$ ,  $i = 1, 2, \dots, N$  and  $N = 1, 2, 3, \dots$ .

These are the time-means, indicated in Section 4 with an accent-- which we drop from now on for simplicity of notation. As in the case of the finite ensemble of waves first treated, we assume that the  $N$  waves in each row  $\psi^N = \{\psi_1^N, \dots, \psi_N^N\}$  are incoherent in phase (our first statistical assumption); and we again denote by  $P(\psi^N)$  the total power they inject into  $K$  across its bounding end  $(\Delta \mathcal{G}, \Delta \Omega)$ . As before, it is a sliding invariant of the rays in  $\mathcal{G}$ .

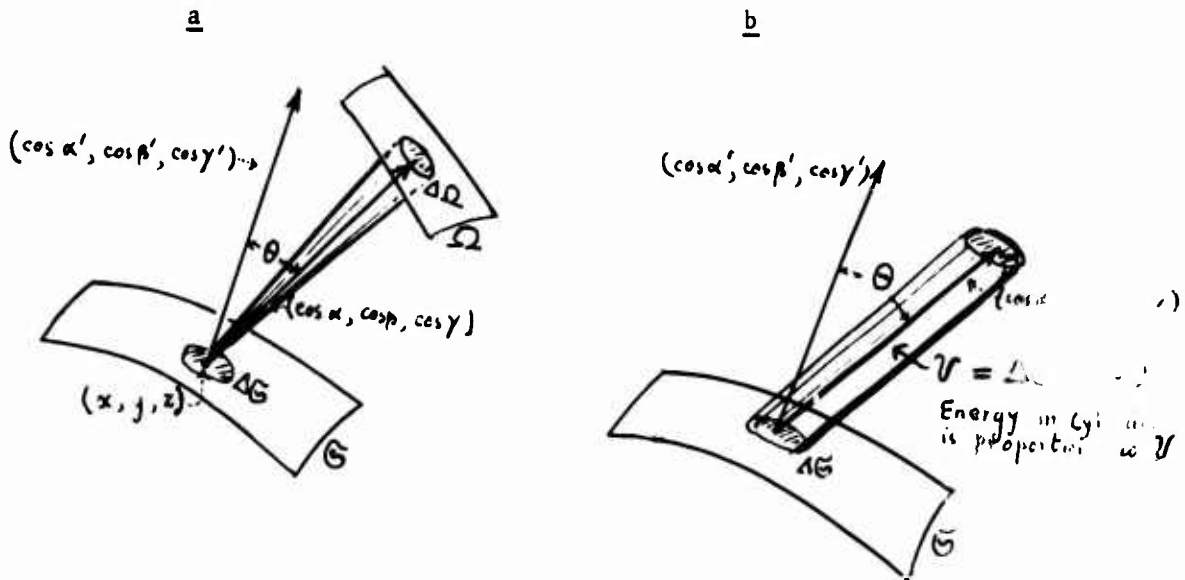
The second part of our statistical assumption is that for every choice of  $(\Delta \mathcal{G}, \Delta \Omega)$  the power  $P(\psi^N)$  approaches a limit  $P(\Psi)$  as  $N$  increases indefinitely. Being a limit of sliding invariants, it too is a sliding invariant; but not yet expressed as an integral.

The third and final part of our statistical assumption enables us to express  $P(\Psi)$  as an integral over the 4-parameter figure  $(\mathcal{G}, \Omega)$  of a power flux density in phase space  $\mathcal{H}$ , with respect to a "4-volume"\* (measured by sums of products of elementary areas on  $\mathcal{G}$  and areas on  $\Omega$ ).

\* Technically stated, an assumption of "absolute continuity" of the limiting  $P(\Psi)$ .

This will make  $P(\Psi)$  an integral invariant (in the more drastic "sliding" sense) and will therefore make the body of theorems governing such invariants applicable to the problems of acoustic transmission.

FIGURE 11.I



The concept of density of flux through  $(G, \Omega)$  involves a limiting process. The geometrical relations are shown in Figure 11.Ia. First, select any point  $(x, y, z)$  on our given surface  $G$  and any direction from this point,  $(\cos \alpha, \cos \beta, \cos \gamma)$ ; these cosines are the coordinates of the representative point on the unit sphere  $\Omega$ . Second, construct a small piece  $\Delta G$  of  $G$  containing this  $(x, y, z)$ , and also a small piece of  $\Delta \Omega$  of  $\Omega$  containing  $(\cos \alpha, \cos \beta, \cos \gamma)$ . The power flow  $\Delta P(\Psi)$  through the elementary 4-dimensional boundary  $(\Delta G, \Delta \Omega)$  is defined as before. The assumption announced above is, firstly, that the ratio  $R$ .

$$R = \frac{\Delta P(\Psi)}{(\text{area } \Delta \mathcal{G}) \cdot (\text{area } \Delta \Omega)}$$

approaches a limit as  $\Delta \mathcal{G}$  and  $\Delta \Omega$  shrink to their respective limiting points,  $(x,y,z)$  and  $(\cos \alpha, \cos \beta, \cos \gamma)$ . Evidently this limit will depend on the six coordinates of these two points, i.e., the point  $(x,y,z)$  and the given direction through it. But it will also depend on the nature and orientation of  $\mathcal{G}$  in the immediate neighborhood of  $(x,y,z)$ . In this latter respect, however, a simple construction familiar in the study of the flow of fluids across surfaces (illustrated in Figure 11.1b) shows that the dependence on  $\mathcal{G}$  reduces to a dependence solely on its positively directed normal at  $(x,y,z)$  (of direction cosines  $\cos \alpha', \cos \beta', \cos \gamma'$ ); and it shows further that the dependence consists merely in the presence of a factor of  $\cos \theta$  in the limit of the above ratio, where  $\theta$  is the angle between this normal and the direction chosen in  $\Omega$ :

$$(11.1) \quad \cos \theta = \cos \alpha \cos \alpha' + \cos \beta \cos \beta' + \cos \gamma \cos \gamma'$$

Therefore the above limit assumption takes the form that, as  $\Delta \mathcal{G} \rightarrow (x,y,z)$  and  $\Delta \Omega \rightarrow (\cos \alpha, \dots)$ ,  $R \rightarrow \epsilon$ , i.e., that

$$(11.2) \quad R \rightarrow \epsilon(x,y,z; \cos \alpha, \cos \beta, \cos \gamma) \cos \theta$$

The assumption of this approach to a limiting density needs to be supplemented by what may seem only a mathematical refinement--the usual

physical assumption of continuity (or piecewise continuity) of this limit as a function of its seven variables, so that its integral over any  $(\mathcal{G}, \Omega)$  is, in fact, equal to power flow across this boundary. With this addition to our third statistical assumption, the statistical description of our radiation field is complete.\* We are able to write the power flow across any  $(\mathcal{G}, \Omega)$  in  $\mathcal{H}$  as the integral

$$\iiint \epsilon \cos \theta \, d\mathcal{G} \, d\Omega,$$

and we know that this is a sliding integral invariant.

It is desirable to re-write the above integral in terms of the canonical coordinates  $(x, y, z, p_x, p_y, p_z)$  in  $\mathcal{H}$ . Applying the relations of (8.5), the first being  $p_x = n(x, y, z) \cos \alpha$ , we see that on the one hand  $\rho$  can be expressed as a function of the canonical coordinates; and on the other hand, that when the direction reference point is on the unit sphere  $\Omega$ , the corresponding momentum vector  $(p_x, p_y, p_z)$  is on a concentric sphere  $\Pi$  of radius  $n = n(x, y, z)$ . Hence, their elements of area are related by the equation  $d\Pi = n^2 d\Omega$  or  $d\Omega = c^2 d\Pi$ . Hence, if we write

$$\begin{aligned} f &= f(x, y, z, p_x, p_y, p_z) \\ &= \epsilon(x, y, z, \cos \alpha, \cos \beta, \cos \gamma) c^2(x, y, z) \end{aligned}$$

we have  $f \cos \theta \, d\Pi = \epsilon \cos \theta \, d\Omega$ , so that our power flow integral becomes

$$(11.3) \quad P(\Psi) = \iiint_{\mathcal{G}} \int_{\Pi} f \cos \theta \, d\mathcal{G} \, d\Pi$$

\* Since so many limiting processes are involved (including  $\omega \rightarrow \infty$ ), assumptions of uniformity are needed. We prefer to regard the limiting steps as forming only an heuristic approach (11.3), the truth of which we postulate in the physical description of a high frequency radiation field.

Acoustic Intensity. The function  $f=f(x,y,z,p_x,p_y,p_z)$  may be termed the flux density of power flow in phase space (due to the ensemble  $\Psi$ ) at the point  $(x,y,z,p_x,p_y,p_z)$  of  $\tilde{\mathcal{S}}$ : the flux density for any orientation of  $\Delta\mathcal{G}$  being  $f\cos\theta$ , it achieves its maximum value when this surface element is so oriented that its normal is in the direction determined by  $(p_x,p_y,p_z)$ , since then  $\theta=0$  and  $\cos\theta=1$ .

If in (11.3) only the integration over  $\Pi$  is carried out, and if this solid angle is chosen as the unit hemisphere having the fixed normal to  $\Delta\mathcal{G}$  as axis (so that  $\theta$  varies between 0 and  $\Pi/2$ ), the following equations are obtained for the total flux of power through the element  $\Delta\mathcal{G}$  in the direction of its normal. Their results are in terms of ordinary xyz-space: the three functions  $U=U(x,y,z)$ ,  $V=V(x,y,z)$ ,  $W=W(x,y,z)$ , and the direction cosines of the normal to the element of surface  $\Delta\mathcal{G}$ . Denoting the result of this operation by  $P(\Psi,\Delta\mathcal{G})$ , the equations are:

$$(11.4) \quad P(\Psi,\Delta\mathcal{G}) = \iint_{\Pi} f \cos\theta \, d\Pi = U \cos\alpha' + V \cos\beta' + W \cos\gamma',$$

in which (11.1) has been used and  $U,V,W$  are given by

$$(11.5) \quad U = \iint_{\Pi} f \cos\alpha \, d\Pi, \quad V = \iint_{\Pi} f \cos\beta \, d\Pi, \quad W = \iint_{\Pi} f \cos\gamma \, d\Pi,$$

and where, in each case,  $f=f(x,y,z,n \cos\alpha, n \cos\beta, n \cos\gamma)$ . Naturally, in carrying out the integrations, one would simplify the limits of integration by a rotation of axes for  $\Pi$  so that the new  $z$  direction (North pole) is parallel to the normal to  $\Delta\mathcal{G}$ ; then the directions belonging to  $\Pi$  are expressed in terms of a co-latitude and longitude, which are

integrated, respectively, from 0 to  $\pi/2$  and from 0 to  $2\pi$ .

Applying (11.4) to the special form of (11.3) when  $\mathcal{G}$  is the element  $\Delta\mathcal{G}$  shows that the total flux of power through this element is given by

$$[\Delta\mathcal{G}] U \cos \alpha' + [\Delta\mathcal{G}] V \cos \beta' + [\Delta\mathcal{G}] W \cos \gamma'$$

where  $[\Delta\mathcal{G}]$  denotes the area of the elementary surface  $\Delta\mathcal{G}$ . Consequently the vector of components  $(U,V,W)$  is the vector flux of power in medium, produced by the radiation field corresponding to the statistical ensemble  $\Psi$ . It is of course the same as the limit of the vector sum of flux vectors  $\vec{F}_i^N$  due to all the elementary waves of  $\Psi$ . The acoustic intensity at the point  $(x,y,z)$  is the length of this vector flux density: Intensity =  $(U^2 + V^2 + W^2)^{1/2}$ . Note that the directional dependence of the phase space flux  $f$  has been averaged out in (11.5): starting with fluxes  $\vec{F}_i^N$  which depend on  $(x,y,z)$  only, we have ended with the total flux due to  $\Psi$ , of components  $(U,V,W)$ , which also depend only on  $(x,y,z)$ . The "mechanism" by which this directional dependence introduced itself into  $f$ ,  $R$ , and  $P(\Psi) = \lim_{N \rightarrow \infty} P(\Psi^N)$ , is of course through the selection process whereby the subset of waves in  $\Psi^N$  was chosen, as well as the way this subset was treated (integrated over  $\Omega$ ), in calculating  $P(\Psi^N)$ .

While this vector field  $(U,V,W)$  in  $xyz$  - space is what is measured and used in practice, its properties and algorithms of calculation are based on those of the phase space flux  $f$ . The detailed study of special cases is given in the succeeding sections; we close this one by proving

a theorem upon which all the applications are based: that  $f$  is constant along each ray in phase space  $\mathcal{H}$ . Therefore, it is a first integral, which, in important special cases may be the trivial one: constant throughout  $\mathcal{H}$ , or in those parts of it taking part in the transmission.

This theorem is a simple consequence of the existence, in addition to the power flow integral invariant  $P(\Psi)$  given by (11.3), of a basic sliding integral invariant common to all conservative Hamiltonian systems.

$$(11.6) \quad I_4 = \iint_{\mathcal{G}} \iint_{\Pi} \cos \theta \, d\mathcal{G} \, d\Pi .$$

As stated toward the close of Section 9, such a 4-dimensional integral invariant can be derived from the 5-dimensional one based on the conservative density  $1/|\nabla H|$ . The process was merely exemplified, using the lower dimensional case illustrated in Figure 9.I. The simple rigorous derivation in Appendix E starts from the universal Hamiltonian "action integral" noted at the end of Section 9. The domain of integration in  $I_4$  can be any 4-dimensional piece of  $\mathcal{H}$ , for example,  $(\mathcal{G}, \Pi)$ .

We thus have two 4-dimensional sliding integral invariants of the system of equations (8.4) in our 5-dimensional phase space  $\mathcal{H}$ , namely  $P(\Psi)$  and  $I_4$ . In view of a general theorem, the ratio of their integrands is constant along every integral curve of (8.4): i.e., it is a first integral of this system (which may be the trivial one: constant throughout  $\mathcal{H}$ ).



This theorem, proved by formulas in Appendix E, can be made intuitively obvious as follows: Let  $C$  be any integral curve; take the "hyper-worm" of neighboring integral curves, containing  $C$ . From the sliding invariance of  $P(\Psi)$  and  $I_4$ , their values over any 4-dimensional cross-cut of this hyper-worm are independent of the cross-cut; hence, the same is true of their ratios. Passing to the limit as the hyper-worm shrivels down to  $C$ , we see that, on the one hand, the limit of the ratio is the ratio of the integrands at the point where this cross cut intersects  $C$ ; and on the other hand, that this ratio is independent of the position of this intersection. Hence, it is constant along  $C$ .

Applied to the present case, we see that the ratio of integrands of  $P(\Psi)$  and  $I_4$ , namely

$$(11.7) \quad \frac{f \cos \theta}{\cos \theta} = f = f(x, y, z, p_x, p_y, p_z)$$

is constant along each integral curve of the Hamiltonian ray equation

(8.4). This is the basis of the Hamiltonian treatment of radiation fields--the whole remainder of the present work.

A remark on the proviso, implied in all these derivations, that  $\theta$  is the angle of our direction with the positively directed normal to  $\mathcal{G}$ : The choice of a direction across  $\mathcal{G}$  is arbitrary; but once made has to be adhered to consistently in interpreting the formulas. If we take it in the direction of flow of power in the propagation, then  $P(\Psi)$  represents the (positive) power actually entering the "hyperworm"  $K$ ; if in the opposite sense, the integral gives the power leaving it (which in such a case would be negative).

## 12. A Method of Statistical Perturbations

The three-fold statistical assumption describing the model of hydro-acoustic radiation, together with its consequences concerning the flow of power along sets of rays, formulated in the preceding section, has led to the theorem that the flux density function  $\mathcal{F} = f(x, y, z, p_x, p_y, p_z)$  is constant along any ray  $L$  in the phase space  $\mathcal{H}$ : it is a first integral of the canonical ray equations (8.4) -- but it may reduce to a constant throughout  $\mathcal{H}$  or certain of its subregions.

From this property of  $f$ , an obvious method can be described for calculating transmission loss in the flow of power from an emitter  $A$  to a receiver  $B$ : Enclose each of these bodies in a surface,  $\mathcal{G}_A$  and  $\mathcal{G}_B$ , the former just containing  $A$  together with the water participating in the hydro-acoustic emission (often taken as a "standard sphere"); the latter so close about  $B$  that all acoustic power crossing it is detected by  $B$ . Since it is usually assumed that the acoustic situation at the emitter is known (or tentatively assumed, in the case of a hostile emitter), we can say that the values of  $f$  are known at all points and directions of  $\mathcal{G}_A$ ; i.e., at all points of the 4-dimensional  $(\mathcal{G}_A, \Pi_A)$ , where  $\Pi_A$  is the set of outwardly oriented  $p$ -vectors from  $\mathcal{G}_A$ . Suppose then that a certain ray  $L$  has been found to cut  $\mathcal{G}_A$  in a direction belonging to  $\Pi_A$ , i.e., in a point  $P_0$  of  $(\mathcal{G}_A, \Pi_A)$ ; and also to cut  $\mathcal{G}_B$  in a direction in  $\Pi_B$  (the set of inward directions to the latter), i.e. at the point  $P_1$  of the 4-dimensional  $(\mathcal{G}_B, \Pi_B)$ . Then since  $f$  is constant along  $L$ , its value at  $P_1$  equals its

value at  $P_0$  which, as we have just explained, is known. Thus, if all the rays which cut  $\mathcal{G}_A$  and  $\mathcal{G}_B$  have been found,  $f$  will be known at all receiver points  $(\mathcal{G}_B, \Pi_B)$  so that the total power received by B from A can be calculated by the formulas of the preceding section.

The picture simplifies greatly if -- as is usually assumed --  $f$  is constant over  $(\mathcal{G}_A, \Pi_A)$ . (Actually the effect on B will be the same as if we only assume  $f$  constant over the smaller set  $(\mathcal{G}_A, \Pi'_A)$ , where  $\Pi'_A$  is the part of  $\Pi_A$  corresponding to directions from  $\mathcal{G}_A$  of those rays that could possibly reach  $\mathcal{G}_B$ ). Thus,  $f$  is constant and known throughout the 5-dimensional "hyper-worm"  $K$  composed of all rays in  $\mathcal{S}$  passing through  $(\mathcal{G}_A, \Pi_A)$ . As frequently stated in earlier sections, the rays can be regarded as the paths of a steady conservative flow in phase space  $\mathcal{S}$ . If, then, the "fluid" becomes stained with ink as it crosses the 4-surface  $(\mathcal{G}_A, \Pi_A)$  in its emission from A, this region  $K$  is the stained part of  $\mathcal{S}$ : it will stain that part  $(\mathcal{G}_B, \Pi_B)_A$  of the receiver surface  $(\mathcal{G}_B, \Pi_B)$  by ink transported in  $K$ . But this  $(\mathcal{G}_B, \Pi_B)_A$  is the part through which flows the power sent from A to B, which power equals the known constant  $f$  multiplied by the "4-dimensional measure" of  $(\mathcal{G}_B, \Pi_B)_A$ , calculated by equation (11.4), the integration being over  $(\mathcal{G}_B, \Pi_B)_A$ . The outstanding difficulty is, of course, in finding the latter region, which requires integrating the ray differential equations (8.4).

The picture will actually be drawn in Section 13, which deals with the "laminar" case  $c = c(z)$  and its extension  $c = c(r, z)$ , in both of which the physical rays are in a vertical plane, the phase space  $\mathcal{S}$  being 3-dimensional.

While the methods just described are theoretically perfect, they become more and more impractical as the range increases, since the mechanical or other computations needed to trace the individual rays become increasingly elaborate. As will be shown in the simplified case in Section 13, with increasing range the rays, and the inked region they sweep out, rapidly become highly contorted after a couple of convergence zones. This could have been expected from an examination of the conventional text book figures traced in quite simple cases (e.g., Fig. 5.19 in Tolstoy and Clay<sup>(1)</sup>). It is the phenomenon of mixing in phase space, which has played such a fundamental role in the ergodic theory (see Appendix G).

What is perhaps an even greater objection to all methods (such as those described above) in which individual rays have to be computed over long ranges (several convergence zones) is their inescapable artificiality: precise values of  $c$  as a function of position have to be assumed before the computations can start; and as every experimental oceanographer knows, all such detailed assumptions are unrealistic. On the other hand, this very difficulty can lead to a simplified point of view which can focus the attention on the realities of the process of transmission loss computation and correspondingly simplify the work: it is the statistical viewpoint and the application of rather simple statistical methods. It turns out that our Hamiltonian picture gives the statistical approach a most natural setting.

Perturbations. Our first step in implementing the ideas of the preceding paragraph is to apply to the Hamiltonian equations of Section 8 the method of perturbations, of fundamental use in so many other branches of

physical science. The idea of this method is simple: the term in the equations due to the external influences (e.g., forces) which in our ray equations is the "potential"  $V$  defined as

$$(12.1) \quad V = V(x,y,z) = -1/2c^2 = -n^2/2,$$

is broken up into the sum of two parts: one part  $U$  is large and simple; the other part  $W$  is complicated but small. The latter is regarded as the "perturbing term". Then a method of successive approximations is applied, starting from a first approximation which is the solution of what the canonical equations (8.4) would become if  $W$  were dropped from the Hamiltonian (8.3),  $V$  being set equal to  $U$ . The method can take many forms; but for our present purpose we shall orient it toward the evaluation of the first order effects of the perturbing function  $W$ , stopping at the second stage and dropping higher powers of the perturbations in the equations.

There are two essentially different situations in the application of the above method, depending on whether  $W$  is known as a definite, although complicated function of  $(x,y,z)$ , or whether it is only known through its statistical properties. The first situation is the classical one of physics and celestial mechanics; it would be applicable to the case of acoustic rays if the acoustic profile changed in a known way with changes in geographical position, e.g., with latitude. The second is the case of our present concern: it requires us to formulate the statistical facts representing our degree of knowledge or ignorance of  $V$ .

The desired statistical model is suggested by the conventional gathering of oceanographic data and predictions bearing on the medium of propagation: the results take the form of averages at different places, depths, and seasons. Accordingly, we shall think of an idealized "population" of conditions, expressed as a statistical ensemble  $\{V\}$  of functions  $V$ , one of which is "drawn at random". The population average -- the expected value  $\bar{V}$  of this function -- shall be used for the given, known function:  $U = \bar{V}$ . In general, the "population average" or expected value of any quantity determined by this statistical ensemble will be denoted by the bar.

It follows from the above choice of  $U$  that  $\bar{W} = 0$ , since  $W = V - U = V - \bar{V}$ . Furthermore, since differentiations and integrations are linear operations and hence interchangeable with this averaging, we have

$$\overline{\partial W / \partial x} = \partial \bar{W} / \partial x = 0, \text{ etc.},$$

It follows also that if  $C$  is any curve in xyz-space and if  $G=G(x,y,z)$  is any given function, the average of any line integral such as  $\int_C G(\partial W / \partial x) ds$  is the line integral of the average:  $\int_C G(\partial \bar{W} / \partial x) \cdot ds$ , which is zero. On the other hand, integrals and other expressions containing non-linear factors, such as squares and products of  $W$  or its partial derivatives, will not necessarily have vanishing averages, except when special assumptions are made. Such special cases, which usually apply to terms containing uncorrelated products, will appear later.

Turning to the Hamiltonian ray equations (8.4), we shall indicate by single accents the first approximation, obtained by replacing  $V(= -n^2/2)$  in the Hamiltonian  $H$  of (8.3) by its average  $U$ . Then the equations (8.4) take the form indicated by the first two that follow (in which  $U' = U(x', y', z')$ ):

$$(12.2) \quad dx'/d\tau = p'_x, \quad dp'_x/d\tau = -\partial U'/\partial x', \text{ etc.}$$

The second approximation, indicated by double accents (and in which we write  $U'' = U(x'', y'', z'')$  and  $W' = W(x', y', z')$ ), is defined as the solution of the six equations, of a type illustrated by the pair

$$(12.3) \quad dx''/d\tau = p''_x, \quad dp''_x/d\tau = -\partial U''/\partial x'' - \partial W'/\partial x'.$$

It is observed that the perturbing term, involving  $W$ , is the derivative of this "random function" calculated for known values of the coordinates; i.e., for the  $x'$ ,  $y'$ ,  $z'$  determined by the first approximation.

We introduce the first order perturbations, defined as

$$(12.4) \quad \begin{aligned} X &= x'' - x', & Y &= y'' - y', & Z &= z'' - z' \\ P_x &= p''_x - p'_x, & P_y &= p''_y - p'_y, & P_z &= p''_z - p'_z \end{aligned}$$

Since we are assuming that the initial values (i.e., at  $\tau = 0$ ) of both approximations are the same (i.e.,  $x'_0 = x''_0$ , etc.), it follows that the perturbations are all zero for  $\tau = 0$ . We must find the differential equations satisfied by the perturbations (to quantities of the first order). For this purpose we first subtract each equation in (12.2) from the corresponding one in (12.3) and apply (12.4), thus obtaining e.g.

$$(12.5) \quad dX/d\tau = P_x$$

$$dP_x/d\tau = -(\partial U''/\partial x'' - \partial U'/\partial x') - \partial W'/\partial x'$$

Next, we note that the first difference on the right is the difference between the known (non-random!) function  $\partial U/\partial x$  calculated for  $(x', y', z')$  and for  $(x'', y'', z'')$ . As we are working only to the first order in the perturbations, we can expand it in a Taylor series about the known (non-random)  $(x', y', z')$ , dropping all powers except the first in the perturbations. The result is

$$(12.6) \quad dP_x/d\tau = U'_{xx} X + U'_{xy} Y + U'_{xz} Z + W'_x$$

in which the  $U'$  with the double subscripts denotes the corresponding second partial derivative, always calculated for the first approximation (single accent) values (which are determined and non-random); while  $W'_x = \partial W'/\partial x'$ .

There results from this process six first order linear differential equations ((12.5), (12.6) and the corresponding pairs for the other two coordinates). The coefficients of the capital letters (random perturbations) are determined functions of  $\tau$  through their containing the first approximation functions  $x'$ , etc., of this independent variable. Since they contain the random perturbing terms  $\partial W'/\partial x'$ , etc., they are not homogeneous.

On taking averages, we find that, since the perturbing terms disappear, the averages  $\bar{X}$ ,  $\bar{P}_x$ , etc., satisfy the corresponding homogeneous



differential equations. Since the quantities are initially all zero, an elementary basic theorem in such equations tells us that they all vanish identically:

$$(12.7) \quad \bar{X} = \bar{Y} = \bar{Z} = \bar{P}_x = \bar{P}_y = \bar{P}_z = 0 .$$

This is, of course, only true to terms of the first order.

In order to investigate the properties of the variances and co-variances, which, in virtue of (12.7) are the means of squares and products,  $\overline{X^2}$ ,  $\overline{XY}$ , etc., we must apply a slightly less elementary theorem in ordinary linear differential equations, the proof of which is outlined in Appendix F in the simple form in which it is needed here. The theorem tells us that we can write our perturbations as integrals of the following forms

$$(12.8) \quad \begin{aligned} X &= \int_0^\tau [G_{xx}(\tau, \xi) W'_x(\xi) + G_{xy}(\tau, \xi) W'_y(\xi) + G_{xz}(\tau, \xi) W'_z(\xi)] d\xi \\ P_x &= \int_0^\tau [K_{xx}(\tau, \xi) W'_x(\xi) + K_{xy}(\tau, \xi) W'_y(\xi) + K_{xz}(\tau, \xi) W'_z(\xi)] d\xi \end{aligned}$$

(along with two more pairs of similar type for  $(Y, P_y)$  and  $(Z, P_z)$ , in which the first of the double subscripts is replaced by y and by z, respectively). In these equations, the functions of  $(\tau, \xi)$  with double subscripts are completely determined by the coefficients of the perturbations  $(X, P_x, \text{etc.})$  in (12.5) and (12.6), which are, in turn, determinate functions, calculable in terms of the known function U. The random effects are produced only by the W-factors. From (12.8), by taking averages, the same result (12.7), obtained before by more elementary methods, would become obvious. But we now can deal with mean squares and products.

We first write  $X^2$  in terms of products of integrals: the expression for  $X$  in (12.8) can be written as the sum of three integrals, so that  $X^2$  is the sum of three squares and three cross products. We shall re-write these with the aid of a simple formula of the calculus, expressing the product of two integrals as the double integral (over a square-shaped region) of the product of the respective integrals, in each of which different variables of integration have been inserted (the coordinates of the point on the square). In its simplest form (for any integrands,  $f$  and  $g$ ) it is

$$\int_0^\tau f(\xi) d\xi \cdot \int_0^\tau g(\eta) d\eta = \iint f(\xi) g(\eta) d\xi d\eta ,$$

the double integral on the right being over the square

$$(\xi, \eta): 0 \leq \xi \leq \tau, 0 \leq \eta \leq \tau.$$

Applying this to such a typical cross-product term in  $X^2$  as

$$\int_0^\tau G_{xx}(\tau, \xi) W'_x(\xi) d\xi \cdot \int_0^\tau G_{xy}(\tau, \eta) W'_y(\eta) d\eta$$

this becomes

$$\iint G_{xx}(\tau, \xi) G_{xy}(\tau, \eta) W'_x(\xi) W'_y(\eta) d\xi d\eta.$$

The mean of this is the integral

$$\iint G_{xx}(\tau, \xi) G_{xy}(\tau, \eta) \overline{W'_x(\xi) W'_y(\eta)} d\xi d\eta.$$

At this point, the first special assumption concerning the ensemble  $\{W\}$  is stated on physical grounds: that the gradients in two perpendicular directions of the random function  $W$  are independent. Therefore

$$\overline{W'_x(\xi) W'_y(\eta)} = \bar{W}'_x(\xi) \cdot \bar{W}'_y(\eta) = 0$$

Hence the mean of this, as well as all other cross-products, vanish.

Turning to a square, we have as before the double integral expression for their means, such as

$$(12.9) \quad \iint G_{xx}(\tau, \xi) G_{xx}(\tau, \eta) \overline{W'_x(\xi) W'_x(\eta)} d\xi d\eta.$$

Here we are dealing with a product of the values of one and the same random function  $W'_x$ , calculated at positions  $\xi$  and  $\eta$  that may coincide or be more or less distant. If they coincide, we have the variance  $\overline{(W'_x)^2(\xi)}$ , obviously a positive quantity. If they are miles apart,  $W'_x(\xi)$  and  $W'_x(\eta)$  will be independent, and the mean in their product will contribute a zero value to the integrand in (12.9). The distance between  $\xi$  and  $\eta$  within which the dependence of  $W'_x(\xi)$  and  $W'_x(\eta)$  becomes appreciable -- which plays the role of a "relaxation interval" -- is not known with any precision to modern oceanography; but it would seem physically natural to assume it not more than a mile or two -- certainly considerably less than one convergence zone. Therefore, only close to the diagonal  $\xi = \eta$  of the square over which the integral in (12.9) is extended will there be any non-zero values of its integrand. Various approximate expressions of (12.9) as an integral along this diagonal could be given on the basis of natural assumptions; but in the incompleteness of our knowledge of the refinements of ocean statistics, the simplest approximation may be the best: to set the dashed factor equal to the variance  $\overline{W_x^2(\xi)}$  time the Dirac delta function,  $\delta(\xi - \eta)$ ; whereupon (12.9) becomes

$$(12.10) \quad \int_0^\tau G_{xx}^2 (\tau, \xi) \overline{W_x'(\xi)^2} d\xi$$

Applying this analysis to the other terms, we obtain (dropping accents and independent variable signs):

$$(12.11) \quad \overline{X^2} = \int_0^\tau [G_{xx}^2 \overline{W_x^2} + G_{xy}^2 \overline{W_y^2} + G_{xz}^2 \overline{W_z^2}] d\xi$$

Similarly

$$(12.12) \quad \overline{XY} = \int_0^\tau [G_{xx} G_{yx} \overline{W_x^2} + G_{xy} G_{yy} \overline{W_y^2} + G_{xz} G_{yz} \overline{W_z^2}] d\xi$$

All the other variances and covariances of the random perturbations can be expressed explicitly by such formulas, using the coefficients in (12.8). The only general facts to retain from such results are, firstly, that the variances in the position along a ray are of the order of the variances in the gradient of the random unknown component  $W$  in the acoustic environment (roughly proportional to them); and increase with the distance (which is roughly proportional to  $\tau$ ). Hence the standard deviation (root mean square) is roughly proportional to the square root of the distance along the ray, and to the standard deviation of the gradient of  $W$ .

This places the validity of detailed computations of rays at long ranges in doubt, although the statistics of acoustic profiles in the ocean will have to be known to a much greater extent than they are at present to say just how far our computed rays suffer how much random deviation. It also suggests that the ink-stained fluid passing across the emitter surface

$\mathcal{G}_A$  may do more than contort itself as in steady streaming: a picture of a more drastic mixing, analogous to turbulence, may be more appropriate. Then the whole of  $\mathcal{G}_B$  (and its directions  $\Pi_B$ ) would be reached by a diluted fluid from A; and the degree of dilution would determine the transmission loss.

These matters, in much more detailed and concrete form, will be taken up in Section 13.

### 13. Rays in a Vertical Plane with Azimuthal Symmetry

In the case of constant acoustic profile, i.e., when the acoustic quantities  $c$ ,  $\rho$ ,  $V$  characterizing the medium depend on depth  $z$  only (the "laminar case"),<sup>\*</sup> the rays are all in vertical planes, and, as shown in Appendix B, their differential equations can be solved by quadratures, expressed in terms of integrals containing these quantities. But as soon as the profile varies (in a more or less known way) along the path, all the complications discussed in the preceding section enter the scene. This is true in spite of the fact that by working in phase space rather than in geographic space, the earlier ambiguities (caustics, multiple-valued wave front functions, etc.) have been banished. The object of the sections 13 - 16 is to take one step beyond the laminar case, with the double regard for achieving a useful degree of realism and for a simplicity in graphical representation (a lower dimensionality than in the general case). This will allow a visual presentation of the ideas and methods of the last two sections, and will lead to simplified numerical methods, useful in themselves.

The facts developed here are derived from four assumptions which specialize the general conditions. As a matter of terminology, we shall call the vertical axis passing through a central reference point A in the emitter the emitter axis, and we shall take it as the  $z$ -axis in a system of cylindrical coordinates. This is shown in Figure 13.IA (for convenience in an "upside down" position, since in the ocean  $z$  increases with depth). Rotation about the emitter axis is measured by the azimuthal angle  $\phi$ ,

---

\* Usually called the case of "layered media".

while horizontal distance from the emitter axis is the range  $r$ . Our first two assumptions of this section are as follows:

1. The acoustic quantities ( $c$ , etc.,) are functions of depth and range only ( $V = V(r, z)$ ).
2.  $V = U(z) + W(r, z)$  and  $W \ll U$ . This  $W$  may be the random perturbation of Sec. 12 ( $\bar{W} = 0$ ); or an irregular departure from laminarity.

Thus in our first approximation we are in the laminar case; range dependence enters only through the random perturbation  $W = V - \bar{V} = W(r, z)$ .

The third and fourth assumptions are based upon a further specialization which will now be explained. We start by picturing the emitter as radiating symmetrically in azimuth: power in all azimuths being the same. This is only an approximation, but deemed good enough to have been used very often. Next, we "backtrack" on our assumption of Section 11 (11.3), that rays through a given point could have a continuum of azimuthal directions: we shall confine them to the emitter's axial plane. Thus the 4-dimensional density  $f$  is replaced by a 3-dimensional  $\tilde{f}$ ; i.e., we set  $f = \tilde{f} \delta(\lambda - \phi)$ , (for  $\vec{p}$ 's azimuth  $\lambda$  and Dirac's  $\delta$ ). Consistently, we assume a corresponding azimuthal symmetry for the bounding surfaces of the ocean. In view of these assumptions, in conjunction with Assumption 1 above, the wave equations (3.1) possesses azimuthally symmetric solutions, since both its coefficients  $c$  and the boundary conditions have this independence of  $\phi$ . Therefore the same is true of the Helmholtz equation (4.2) and consequently of the elementary travelling waves. Such pairs as (A, B) being independent of azimuth  $\phi$ , it follows by (4.3) and (4.4) that the power flow lines will lie in co-axial planes (all containing the emitter axis). Furthermore, the power flux vectors will have the same azimuthal symmetry. Obviously not all solutions of our azimuthally symmetric equations and

symmetric boundaries will have this symmetry: our statement is, simply, that such symmetrical solutions exist, and exist in sufficient numbers so that we can, without mathematical inconsistency, construct our ensemble  $\Psi$  out of them exclusively. Our third and fourth assumptions are consequences of selecting in the statistical ensemble  $\Psi$  of Section 11 only such solutions  $\psi_1^N$  which are functions of  $r$  and  $z$  only.

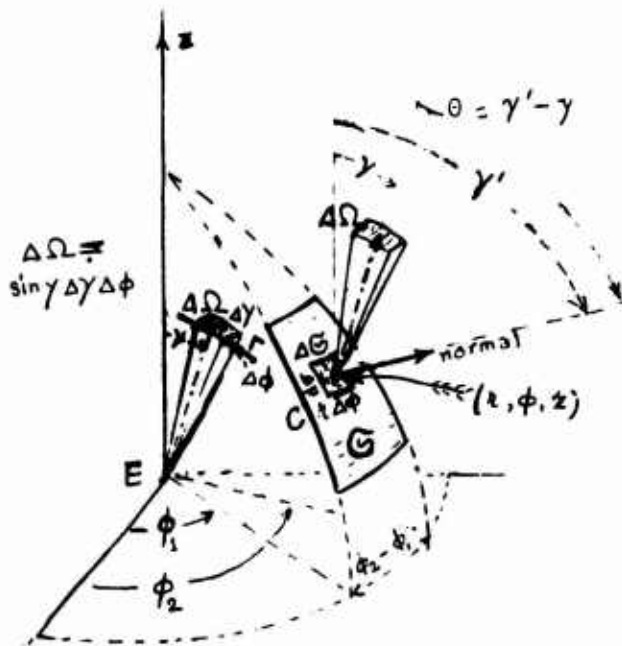


Figure 13.IA. Element  $(\Delta G, \Delta \Omega)$  in Cylindrical Coordinates.

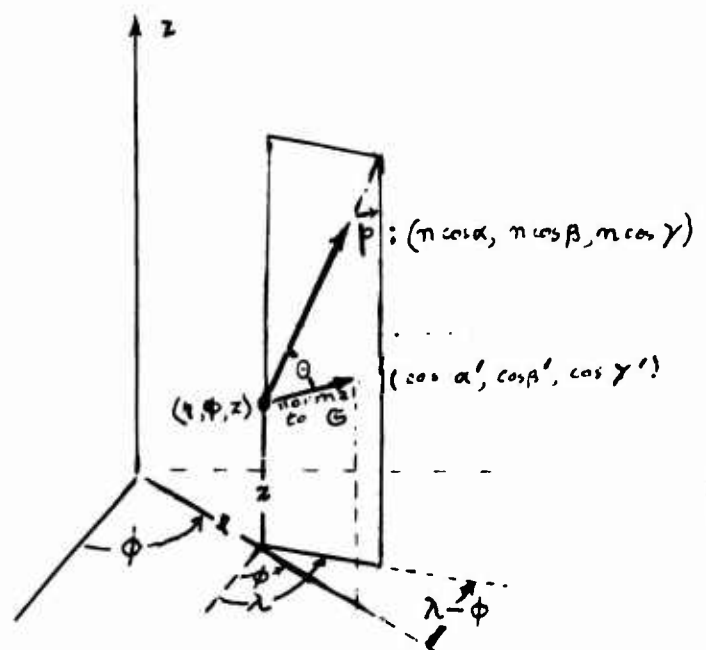


Figure 13.IB. Momentum Vector  $\vec{p}$  attached to  $(r, \phi, z)$  in 13.IA.

To explore the consequences of the azimuthal symmetry in the present situation, we recall the assumptions of Section 11 concerning the ensemble  $\Psi$ , formulated in (11.2) and (11.3), and depicted in Figure 11.I. We shall make an adaptation to the present case of the construction of the surface  $\mathcal{G}$  and solid angle  $\Omega$ , requiring both to be figures of revolution about the  $z$ -axis through the common angle  $(\phi_1 \leq \phi \leq \phi_2)$ , the former generated by a curve  $C$  traced in one of the co-axial planes; the latter, generated by the

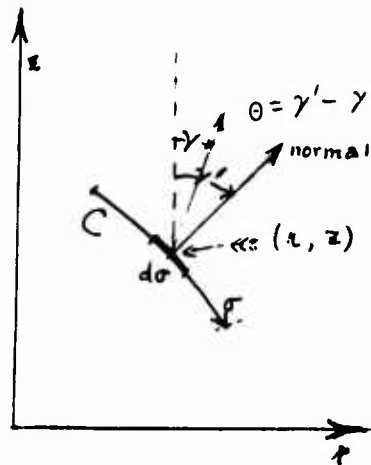


arc  $\Gamma$  of a unit circle in the same co-axial plane, but centered at a point on the axis. Similarly, the small pieces  $\Delta\mathcal{C}$  and  $\Delta\Omega$  are generated by the revolution, through a common small sub-angle  $\Delta\phi$ , of the small pieces  $\Delta C$  and  $\Delta\Gamma$  of  $C$  and  $\Gamma$ . The construction is shown in Figure 13.1A but with  $\Omega$  and  $\Delta\Omega$  transported so that their vertex is at a point  $(r, \phi, z)$  on  $\Delta\mathcal{C}$ , in analogy with Figure 11.1. Evidently (to quantities of higher order)  $\Delta\mathcal{C} = r\Delta\phi\Delta C$  and  $\Delta\Omega = \sin\gamma\Delta\Gamma$ . Therefore, in surface integration over  $(\mathcal{C}, \Omega)$ , using the arc-length  $\sigma$  along  $C$  (in a given sense) and the angle  $\gamma$  (with the  $+z$ -axis) along  $\Gamma$ , we have

$$(13.1) \quad d\mathcal{C} = r d\phi d\sigma, \quad d\Omega = \sin\gamma d\gamma d\phi$$

The equations (11.2) and (11.3) still hold in the  $\delta$ -function interpretation. Since all elements in the construction have azimuthal symmetry, the same will be true of the power flux quantity  $\tilde{f}$ . Also, the angle  $\Theta$  between the normal to  $\mathcal{C}$  (at the limiting point of  $\Delta\mathcal{C}$ ) and the limiting direction of  $\Delta\Omega$ , is simply  $\Theta = \gamma' - \gamma$ , where as in (11.1),  $\gamma'$  and  $\gamma$  are the angles that the normal and the limiting direction make with the  $+z$ -axis. Finally, since no power flow line in the present case has any azimuthal component, power emitted in directions between any pair of co-axial planes will always remain between them. Putting these facts together, we have a basis for our last two assumptions:

3. The power flux density  $\tilde{f}$  in phase space is a function of range  $r$ , depth  $z$ , and angle  $\gamma$  only; and  $f = \tilde{f}\delta(\lambda - \phi)$ .
4. No power crosses any co-axial plane  $\phi = \text{constant}$ .



**Figure 13.II** Coordinates in Typical Vertical Plane  $\phi = \phi^0$

The Hamiltonian in our cylindrical coordinates is

$$(13.2) \quad H = (1/2)[(p_r^2 + p_\phi^2/r^2 + p_z^2)] - V, \quad V = -n^2/2.$$

The canonical equations of the rays are

$$(13.3) \quad \begin{cases} dr/d\tau = p_r, & d\phi/d\tau = p_\phi/r^2, & dz/d\tau = p_z \\ dp_r/d\tau = -\partial W/\partial r, & dp_\phi/d\tau = 0, & dz/d\tau = -\partial V/\partial z \end{cases}$$

[Of course when the new independent variable  $t$  is introduced through the equation  $dt = n^2 d\tau$  and the momenta are eliminated, (13.3) reduces to the form (B.2) of equations in cylindrical coordinates.]

The fifth equation in (13.3) shows that  $p_\phi = p_\phi^0$ , a constant; the second, that  $p_\phi^0 = r^2(d\phi/d\tau)$ . Since we are only considering rays which cut the  $z$ -axis, where  $r = 0$ , we must have  $p_\phi^0 = 0$ ; therefore, again by the second equation,  $d\phi/d\tau = 0$ , so that  $\phi = \phi^0$ , a constant. This fact, that all the rays considered are in vertical planes, co-axial with the  $z$ -axis, is also derived in Appendix B. On the other hand, Snell's law (B.3)

or, in the present case,  $p_r = k$  (the Snell constant) is only true when  $\partial W/\partial r = 0$ , as we see from the fourth equation. Now since we are regarding the perturbing term  $W$  as giving the departure from laminarity,  $W$  will depend on  $r$ , and hence Snell's law is valid only in the first approximation (i.e., when  $W$  is replaced by zero).\*

Returning to the Hamiltonian system, (13.2) and (13.3) reduce, in view of the fact established above that  $p_\phi = 0$  and  $\phi = \phi^0$ , to the form

$$(13.4) \quad H = \frac{1}{2} [p_r^2 + p_z^2 - n^2]$$

$$(13.5) \quad \begin{cases} dr/d\tau = p_r, & dz/d\tau = p_z ; \\ dp_r/d\tau = -\partial W/\partial r, & dp_z/d\tau = -\partial V/\partial z = -\partial U/\partial z - \partial W/\partial z . \end{cases}$$

These are the equations of a particle moving in a plane of rectangular coordinates  $(r, z)$ , under the action of forces derived from the potential  $V = V(r, z)$ . As in Section 8, our rays are those solutions of (13.4) in the space of the variables  $(r, z, p_r, p_z)$  for which the constant value of our present Hamiltonian (13.4) is zero. Thus our present phase space  $\mathcal{H}$  is 3-dimensional: we shall denote it by  $\mathcal{H}_3$  to distinguish it from the one in the 5-dimensional case. Finally, in terms of the arc length  $s$  along the ray and the relation  $ds = n d\tau$ , the first two equations (13.5) show that

$$(13.6) \quad p_r = ndr/ds = n \sin \gamma, \quad p_z = ndz/ds = n \cos \gamma,$$

which correspond with (8.5), and give the trigonometric reason for the equation  $H = 0$ .

---

\* Despite a too common tendency to regard Snell's law as having the same degree of generality as Fermat's principle. Cf. Appendix A.

The "action integral" along the curve C (c.f. Section 9, end) becomes in the present case

$$I(C) = \int_0 p_r \delta r + p_z \delta z$$

On applying the general methods of Appendix E, this yields the 2-dimensional sliding integral invariant which (in the notation of that appendix) is

$$(13.7) \quad I_2 = \iint (\delta p_r \wedge \delta r + \delta p_z \wedge \delta z)$$

and which, with the choice of region of integration depicted in Figure 13I becomes

$$(13.8) \quad I_2 = \int_C \int_{\Gamma} n \cos \theta \, d\sigma d\gamma$$

This  $I_2$  takes the place of the  $I_4$  of (11.6). Here C is any curve in the rz-plane,  $d\sigma$  an element of arc-length along it,  $\gamma'$  the angle from the +Z-axis to (the positively oriented) normal to C,  $\theta = \gamma' - \gamma$ , the angle between it and the direction of a ray. Finally,  $\Gamma$  is any angular interval (arc of the unit circle). See Figure 13.II.

We must now express the power flux P in terms comparable to (13.8). We first recall equation (11.3), to which we apply the special choice of  $\mathcal{G}$  and  $\Omega$  of Figure 13I, and the expressions (13.1) for  $d\mathcal{G}$  and  $d\Omega = n^2 d\Omega$ . We are then able to write (11.3) as the iterated integral

$$P(\Psi) = \int_{\phi_1}^{\phi_2} d\phi \int_C r \, d\sigma \int_{\phi_1}^{\phi_2} d\lambda \int_{\Gamma} \sin \gamma \, d\gamma \cdot n^2 \tilde{f} \delta(\lambda - \phi) \cos \theta$$

[ $\phi$  and  $\lambda$  are two variables of integration, the first on  $\mathcal{G}$ , the second on  $\Omega$ ]. Now apply our Assumptions 1 and 3. It is necessary to use the cylindric coordinate expression for the direction cosines of the normal and the momentum. These are obtained by projecting these vectors on the horizontal

plane and then projecting the results on the two horizontal axes. By trigonometry, we obtain for the normal

$$\cos \alpha' = \sin \gamma' \cos \phi, \cos \beta' = \sin \gamma' \sin \phi,$$

and similarly, with the azimuth  $\lambda$  of a general momentum vector, illustrated in Figure 13.IB,

$$\cos \alpha = \sin \gamma \cos \lambda, \cos \beta = \sin \gamma \sin \lambda$$

Consequently, after a trigonometric reduction,

$$\begin{aligned} \cos \Theta &= \cos \alpha' \cos \alpha + \cos \beta' \cos \beta + \cos \gamma' \cos \gamma \\ &= \cos(\lambda - \phi) \sin \gamma' \sin \gamma + \cos \gamma' \cos \gamma \end{aligned}$$

Inserting this into the above expression for  $P(\Psi)$ , we note that the  $\lambda$ -integration, using the properties of the  $\delta$  function, converts  $\cos(\lambda - \phi)$  into  $\cos 0 = 1$ ; while the subsequent  $\phi$ -integration merely multiplies the resulting  $\phi$ -independent integrand by the constant factor  $\phi_2 - \phi_1$ ; and we have, with  $\Theta = \gamma' - \gamma$ ,

$$(13.9) \quad P(\Psi) = (\phi_2 - \phi_1) \int_C d\sigma \int_\Gamma n^2 r \tilde{f} \cos \Theta \sin \gamma d\gamma.$$

As the 2-dimensional domain of integration  $(C, \Gamma)$  is slid along the rays in our phase space  $\tilde{\Omega}_3$  the power  $P(\Psi)$  and the angular factor  $(\phi_2 - \phi_1)$  remain unchanged. Therefore the coefficient of  $(\phi_2 - \phi_1)$ , the power flux per unit azimuth angle,

$$(13.10) \quad P_2 = \int_C \int_\Gamma n^2 r \tilde{f} \cos \Theta \sin \gamma d\sigma d\gamma$$

is a sliding integral invariant, and hence, by the theorem cited in Section 11, the ratio of its integrand to that of  $I_2$  is constant along the rays in  $\tilde{\Omega}_3$ ; thus we have, using (13.6),

$$(13.11) \quad n \cdot \sin \gamma \cdot r \cdot \tilde{f} = p_r \cdot r \cdot \tilde{f} = p_r \cdot r \cdot \tilde{f}(r, z, p_z) = \underline{\text{const.}} (\text{on each ray}).$$

It may of course be constant throughout all or part of  $\tilde{\Omega}_3$ . In any case, for a ray issuing from the point  $(r_0, z_0)$  close to A in a direction at an angle  $\gamma_0$ , (so  $p_z = p_z^0 = n(r_0, z_0) \cos \gamma_0$ ), we have

$$n(r, z) \sin \gamma \cdot r \tilde{f}(r, z, p_z) = n_0 \sin \gamma_0 \cdot r_0 \tilde{f}_0$$

Hence

$$(13.12) \quad \tilde{f}(r, z, p_z) = \frac{r_0}{r} \cdot \frac{n_0 \sin \gamma_0}{n \sin \gamma} \cdot \tilde{f}_0 = \frac{r_0}{r} \cdot \frac{p_r^0}{p_r} \cdot \tilde{f}_0$$

The presence of the quotient  $r_0/r$  on the right is what produces the effect of "cylindrical spreading".

It is emphasized that all these results have been derived without assuming laminarity.<sup>\*</sup> Moreover, the assumptions which have led to them can be weakened, by postulating only that azimuthal symmetry applies when  $\phi$  varies through a limited angular range: the range just including the rays that could possibly reach the receiver. Thus the results can be applied to cases of some dissymmetry of emission.

In the case of laminarity -- e.g., in the first approximation with  $W = 0$  -- Snell's Law applies, giving the result  $p_r = p_r^0 = k$ ; then (13.12) reduces to

$$(13.13) \quad \tilde{f}(r, z, p_z) = \frac{r_0}{r} \tilde{f}_0.$$

---

\* Nor, indeed, making use of the approximate laminarity expressed in Assumption 2.

Returning to the more general case, the fact that  $p_r$  is always close to  $p_r^0$  makes (13.13) a good approximation. To calculate the acoustic intensity, equations (11.4) and (11.5) and the further considerations of Section 11 take on a somewhat simpler form in the case of azimuthal symmetry and the cylindrical coordinates. The same is true of the perturbation methods of Section 12. Thus in the cylindrical equivalents of (12.8) there are only two terms in the integrand, arising from the two non-vanishing derivatives of  $W$ . There are corresponding simplifications in the formulas corresponding to (12.11) and (12.12) derived from the earlier one. But basically it is not these various simplifications that are the important results of the azimuthal symmetry: the really useful consequences come from what can be inferred from the mixing and random mixing in the phase space  $\mathcal{H}_3$  as we can <sup>now</sup> visualize it.

These consequences will be set forth graphically in the next Section. We can say in anticipation that there are intermediate range cases in which the constant in (13.11) is a function of  $p_r$  only, so that  $rf$  has this property, thus bringing far greater simplifications than any of the preceding paragraph. At still greater ranges, this quantity is a constant throughout a larger region of  $\mathcal{H}_3$ , with still greater simplifications. The derivation and use of these facts are most easily made with the use of a graphical device for representing the power relations, the surface of section.

#### 14. Graphical Representations and the Surface of Section

This section continues the study of the azimuthally symmetric propagation, with acoustic quantities depending on depth and range only and rays in vertical planes, of Section 13. It takes the mathematical results established there and puts them into graphical form, so that their inter-relations can be visualized. In addition to clarification on the conceptual level, the graphical presentation is shown to lead to simple practical methods of predicting transmission loss at long and at intermediate ranges. Among other things, the effects of such underwater obstructions as sea mounts and ridges can be examined and estimated graphically.

The pictorial representation. This graphical method is made possible by the 3-dimensionality of our phase space  $\mathcal{H}_3$ , in contrast to the 5-dimensionality of  $\mathcal{H}$  in the general case. Our present  $\mathcal{H}_3$  is in fact the locus of "points"  $(r, z, p_r, p_z)$  satisfying the equation  $p_r^2 + p_z^2 = n^2(r, z)$ . To visualize the relationships in this manifold, we must represent its points in our ordinary 3-dimensional space, and this is naturally done by means of a coordinate system. In Figures 14.I and 14.II, which show the typical vertical plane through the emitter axis ( $\phi = \phi^0$ , constant) of the rays in ordinary geographical space, the coordinates are  $(r, z)$  and they are restricted by the three conditions now to be given. Firstly,  $r$  is positive. Secondly,  $z \geq z(\text{surface}) = a$ . This constant surface value  $a$  of  $z$  may be positive, negative, or zero, depending on our choice of the origin of the cylindrical coordinates: we may take a point on the ocean surface (then  $a = 0$ ), at the emitter



A, or at the center of an acoustic duct (minimum  $c$ ); in either case,  $a < 0$ . An origin above the surface would lead to  $a > 0$ ; but this is seldom used. Thirdly,  $z \leq z(\text{bottom}) = b$ . This is an obvious condition; but only if the bottom is assumed to be flat do we have  $b = \text{constant}$ ; otherwise,  $b = b(r)$ , a function of range, corresponding with variations of the ocean bottom along the direction of propagation in our vertical plane. We shall, accordingly, take for our first two coordinates in the representation of  $\mathcal{S}_3$  the  $r$  and  $z$  restricted as follows

$$(14.1) \quad r \geq 0; \quad a \leq z \leq b(r).$$

The third coordinate in the representation of  $\mathcal{S}_3$  must, at each given  $(r, z)$ , specify the momenta  $p_r$  and  $p_z$ ; and vice versa. Since our  $\mathcal{S}_3$  is made up of rays which are paths of power leaving the emitter axis, not approaching it, their tangents are directed away from the  $z$ -axis, so that  $0 \leq \gamma \leq \pi$ . Evidently when  $(r, z)$  are given, each value of  $\gamma$  in this interval determines  $p_r$  and  $p_z$ , in view of (13.6); and vice versa, each pair of values of the latter determines a unique value of  $\gamma$  in the above interval. Consequently the angle  $\gamma$  could be used as the third coordinate in the specification of points in  $\mathcal{S}_3$ . For some purposes  $\gamma$  (or equivalently the angle with the horizontal,  $(\alpha = \pi/2 - \gamma)$ ) is preferable; but for our present investigations simpler formulas and graphs are obtained by using the momentum component  $p_z = n(r, z) \cos \gamma$ . The possibility of this choice is due to the fact that as  $\gamma$  increases from 0 to  $\pi/2$  and thence to  $\pi$ ,  $p_z$  will decrease from  $n$  to 0 and thence to  $-n$ , so that the intervals  $(0 \leq \gamma \leq \pi)$  and  $(-n \leq p_z \leq n)$  are mapped in a continuous one-to-one correspondence, although with reversal of direction.

Note that this would not have been true if  $p_r$  had been used. We shall accordingly represent the points of  $\tilde{\Omega}_3$  by values of  $(r, z, p_z)$ , satisfying (14.1) and also

$$(14.2) \quad -n(r, z) \leq p_z \leq n(r, z).$$

Graphically, inequalities (14.1) and (14.2) determine a solid in the space of the rectangular coordinates  $(r, z, p_z)$ : this is our representation of  $\tilde{\Omega}_3$ . It is drawn schematically in Figure 14.I for the laminar case  $n = n(z)$  and in Figure 14.II for the more general case  $n = n(z, r)$ . Since in the former, the lateral boundaries of  $\tilde{\Omega}_3$  are the loci of the equations  $p_z = \pm n(z)$  (not containing  $r$ ) they are cylindrical surfaces whose elements are parallel to the  $r$ -axis. Furthermore, the locus of points in  $\tilde{\Omega}_3$  for which  $p_r = k$  (i.e. of given Snell constant) is made up of rays in the laminar case, since  $p_r$  is then a first integral of the Hamiltonian ray equations. In terms of our coordinates  $(r, z, p_z)$ , this locus has the equation (derived from (13.6) and  $H = 0$ )

$$(14.3) \quad n(z)^2 - p_z^2 = k^2.$$

Again the coordinate  $r$  is absent from the equation, whose locus is therefore a cylindrical surface of elements parallel to the  $r$ -axis. Such a cylinder is shown in Figure 14.I, with a ray having the corresponding value of  $k$  and therefore lying on this surface. The geometrical interpretation of the variables (13.6) shows that this curve must wind around the surface in a helix-like manner -- and with negative screw rotation as  $r$  increases.

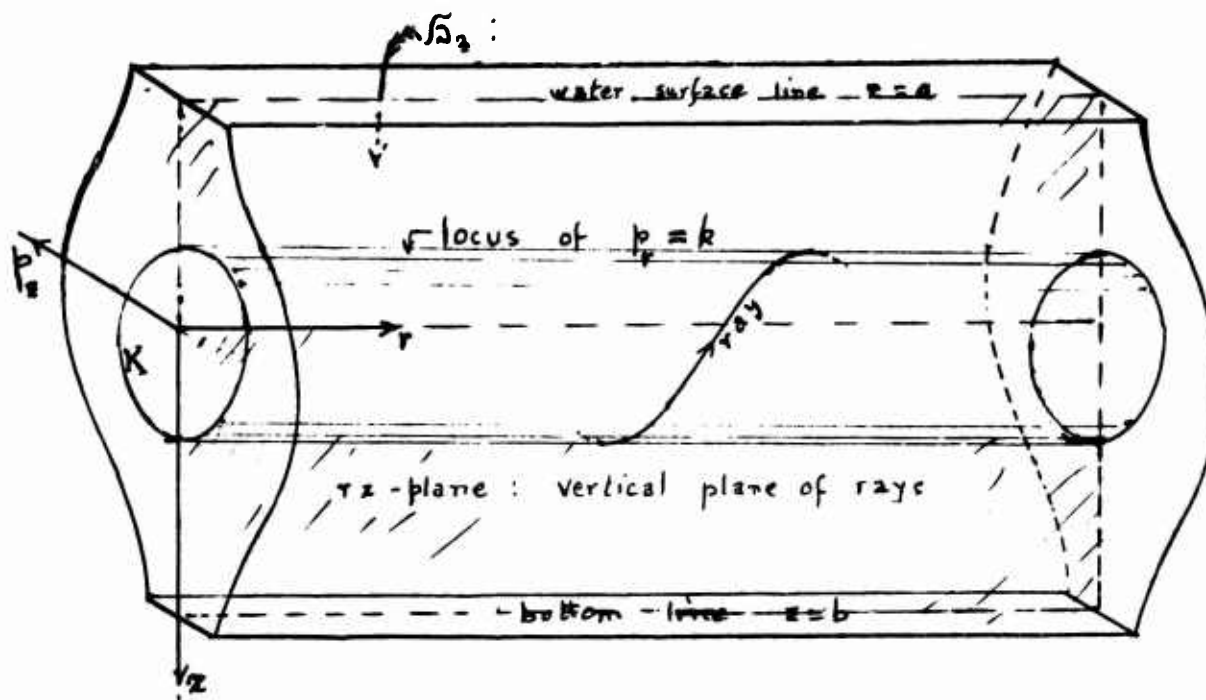


FIGURE 14.I PHASE SPACE IN LAMINAR CASE

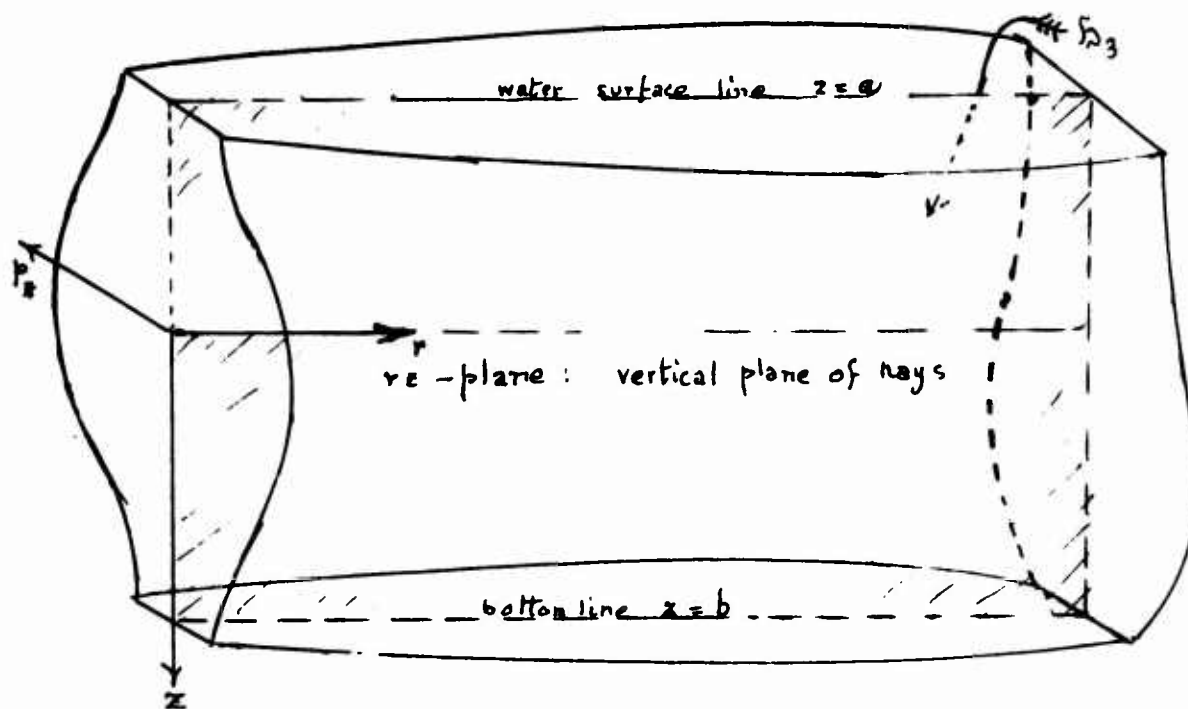


FIGURE 14.II PHASE SPACE IN GENERAL CASE

In the general case shown in Figure 14.II the lateral boundaries of  $\mathcal{H}_3$  are the loci of  $p_z = \pm n(r,z)$ , and are not cylindrical -- except approximately when  $n$  varies only slightly with  $r$ : the "quasi-laminar" case. As was noted in Section 13,  $p_r$  is not constant along each ray; therefore the locus  $n^2 - p_z^2 = k^2$  is not only not a cylindrical surface, but is not generated by rays: a ray having a point on it will in general pass through it, cutting it at an angle. This angle will be close to zero -- sometimes greater, sometimes less -- in the quasi-laminar case. However, in all cases the lateral boundaries are never crossed by the rays involved in propagation over appreciable distances, since they correspond to points where a ray is directed downward ( $\gamma = 0$ ) or upward ( $\gamma = \pi$ ). In the laminar case these boundaries are made of rays that bounce up and down along the vertical (to be rapidly absorbed, of course).

The ocean boundaries. In all cases, any ray that meets the horizontal boundary  $z = a$  (the ocean surface) will experience a certain degree of reflection, absorption, and scattering. These effects are particularly complicated at short ranges, when the angles made with the surface are large, and special physical assumptions are required for their treatment. On the other hand, only nearly horizontal rays ( $\gamma$  close to  $\pi/2$ , so that  $p_z$  is close to 0) can be effective in longer range propagation (many convergence zones): the assumption of specular reflection (with phase shift) has been found to be consistent with the physical observations. We shall make this assumption of loss-less specular reflection at the ocean surface for long and for intermediate ranges. In consequence, any ray which cuts the surface at an angle

$\gamma(>\pi/2)$  is reflected so that  $\gamma$  is replaced by  $\pi-\gamma$  ( $<\pi/2$ ). Therefore the point on the ray where it arrives at the surface,  $(r, a, p_z < 0)$ , jumps to the point where it departs,  $(r, a, p'_z = -p_z > 0)$ . This discontinuity (break in the ray) as it is pictured in  $\tilde{N}_3$  corresponds to a sudden change in angle in the geographic space of the  $rz$ -plane.

The behavior at the ocean bottom is still more complicated, particularly at short ranges. Each encounter involves a complex mixture of absorption, and reflection in directions and amounts depending on the physical nature as well as the shape and depth of the bottom. Research on the subject ranges all the way from geological echo soundings, the extent of bottom bounce (so important in short range detection with certain types of bottom), through phenomena at considerable ranges, when the repeated loss of power with each reflection takes out all that is propagated. This loss seems to be much greater under these circumstances than is the case for surface reflection, and the simplifying assumption that for long and intermediate ranges the cumulative absorption is total has usually shown itself to be justified and will be made here. The  $\gamma$  of the ray through  $(r, z)$  just touching bottom is the critical angle.

The surface of section. The whole state of affairs described here can be given a graphical and kinematic interpretation as follows: First, we cut the solid representing  $\tilde{N}_3$  in Figures 14.I and 14.II by a vertical plane perpendicular to the  $r$ -axis (the locus of  $r = \text{constant}$ ). The resulting figure is a plane region  $\Sigma$  (or  $\Sigma(r)$ ), bounded horizontally by the condition  $a \leq z \leq b(r)$ , and laterally by  $-n(r, z) \leq p_z \leq n(r, z)$ , as is clear from (14.1 and (14.2). This  $\Sigma$  will be independent of  $r$  in

the laminar case (except when possible variations of depth  $b$  are considered); but it will vary more or less as  $r$  is changed in the non-laminar one. In all cases,  $(z, p_z)$  serve as coordinates of points in  $\Sigma$ .

Every ray will cut  $\Sigma$  in one and only one point as it traverses this surface with increasing  $r$ . (The only exceptions occur for vertical rays ( $\gamma = 0, \pi$ ) in the laminar case). This intersection will be called the ray's representative point. Now let the vertical plane  $\Sigma(r)$  be moved continuously to the right -- i.e., let  $r$  run continuously through increasing values --: the representative point on each ray will move continuously in  $\Sigma$ , except if it encounters the upper boundary  $z = a$ , when it will jump to the corresponding symmetrical point by a reversal of sign of  $p_z$ , corresponding with surface reflection. Another exception occurs when a ray meets the lower boundary of  $\Sigma$ :  $z = b(r)$ . Since we are assuming total absorption at such encounters, we shall simply terminate such rays. Thus we have defined a "flow" or continuous one-parameter family of transformations of  $\Sigma$  on itself (more explicitly, of  $\Sigma(r_0)$  onto  $\Sigma(r_1)$ ), with the exceptional situation just noted on its boundaries.

This transformation of  $\Sigma$  is area-preserving

To prove this theorem, let  $A_0$  be a plane region within  $\Sigma(r_0)$  bounded by the curve  $C_0$ . As  $r$  increases from  $r_0$  to  $r_1$ , the points of  $A_0$  and  $C_0$  will move into points of new figures  $A_1$  and  $C_1$ : by construction the latter result from a sliding of the former along rays in  $\Sigma$ . Therefore the "phase integral"  $I(C)$  of Section 13, leading to (13.7),

has the same values:  $I(C_0) = I(C_1)$ . But on these surfaces  $r$  is constant, so  $\delta r = 0$ ; hence the equation becomes

$$\int_{C_0} p_z \delta z = \int_{C_1} p_z \delta z.$$

But these integrals are the negatives of the areas of  $A_0$  and  $A_1$ , as is obvious by the geometrical interpretation of each line integral (cf. the case of  $\int_C y \delta x$ ); or alternatively, by application of Green's theorem in the plane. We could have obtained the same result from the sliding invariance of  $I_2$ : on the plane  $r = r_0$ , (13.7) shows that  $I_2 = \iint_{A_1} \delta p_z \wedge \delta z$ , and the integrand is, in the notation of Appendix E, the element of area in the  $zp_z$ -plane (apart from sign), usually denoted by  $dz dp_z$ . Thus from the invariance of  $I_2$  we obtain the invariance of the area in question. A proof starting with (13.8) and the fact that on  $\Sigma$ ,  $\theta = \pi/2 - \gamma$ , can also be given - but reduces essentially to the one above. It has been assumed in this proof that the rays connecting  $A_0$  and  $A_1$  nowhere meet the bottom; this is essential. On the other hand, the result can be extended to specular reflection of rays at the water surface by an easy construction based on the symmetry in  $z$ -axis.

The surface  $\Sigma = \Sigma(r)$  and the area-preserving continuous transformation -- or "flow" -- induced in it by the rays in  $\tilde{\Omega}_3$  is the application to the present case of the concept of the surface of section and its transformation, introduced into Hamiltonian dynamics by H. Poincaré and his followers at the turn of the century. See Poincaré,<sup>5</sup> Birkhoff.<sup>6</sup>

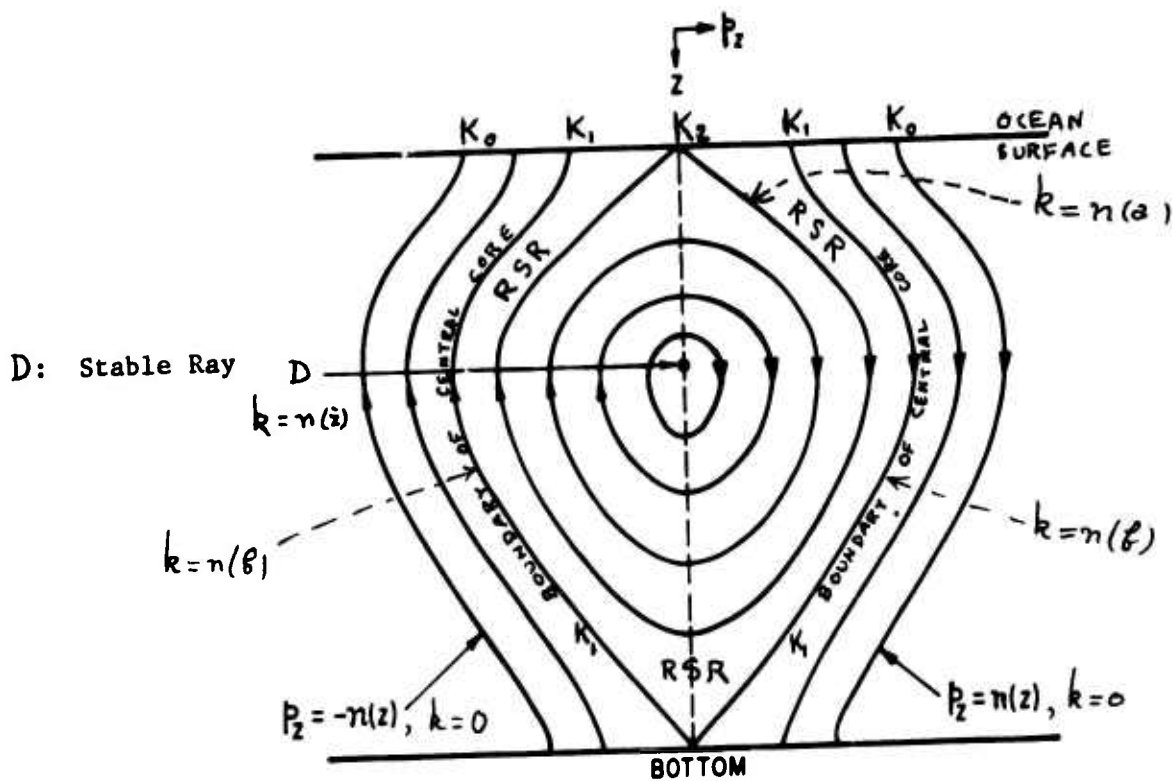


Figure 14.III - Single Duct

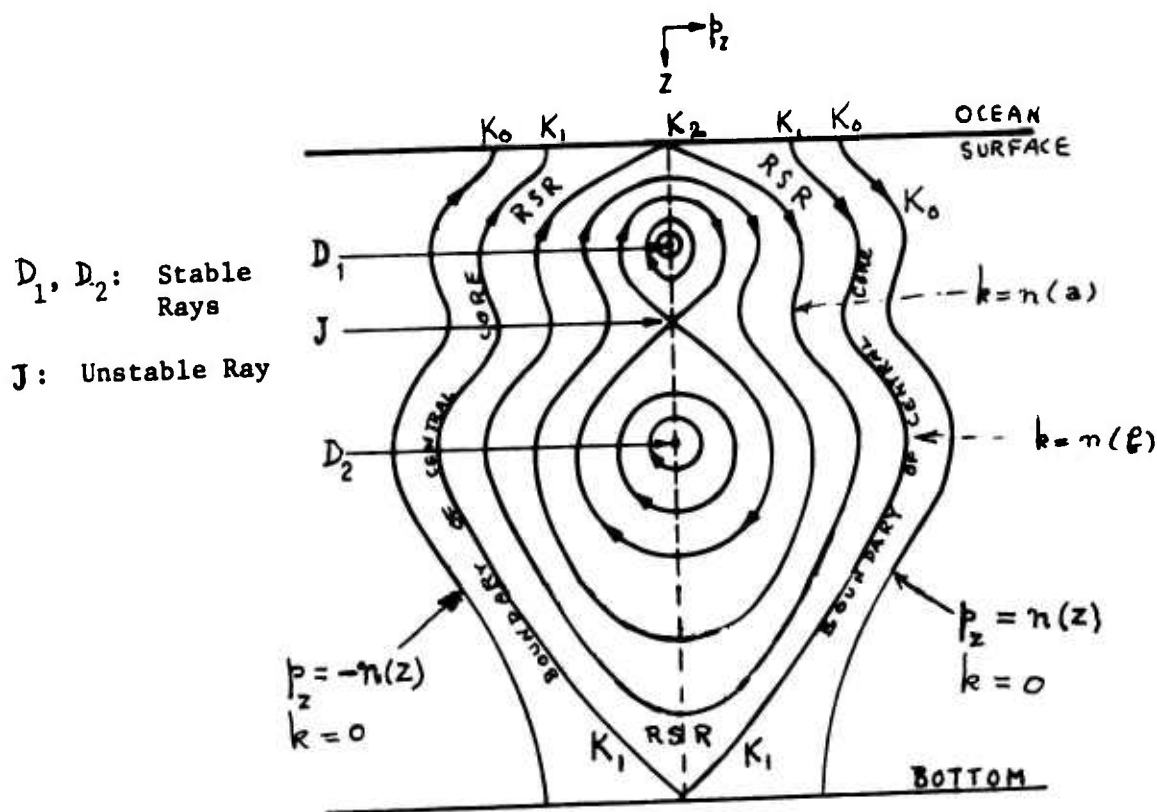


Figure 14.IV - Double Duct



In the laminar case, the shape of  $\Sigma(r)$  is independent of  $r$ .

During the increase of  $r$ , the flow in  $\Sigma(r)$  is along curves  $K$  of given Snell constant  $k$ : they are the orthogonal projections onto  $\Sigma(r)$  of the helix-like curves of the sort shown in Figure 14.I. The variety of possible paths in the laminar case is shown in Figure 14.III when there is a single duct (sound speed minimum) and in Figure 14.IV for two ducts. These correspond to stable horizontal rays, cutting  $\Sigma$  in the point  $D$  and the two points  $D_1$  and  $D_2$ . Between the latter in Figure 14.IV is a point  $J$  where the sound speed has a relative maximum, giving rise to an unstable horizontal ray, whose representative point  $J$  is the multiple point of the locus of points having the same value of  $k$ ; it has the form of a "figure eight". The representative point of any other ray with this same value of  $k$  moves along the figure eight, either approaching -- but never reaching --  $J$  as  $r$  increased indefinitely; or else moving away from it, taking a longer and longer "time" ( $r$ ) as its starting point is taken closer and closer to  $J$ . This situation becomes geometrically evident when the cylindrical surface having this figure eight as directrix and elements parallel to the  $r$ -axis is drawn in  $\mathbb{S}_3$  as in Figure 14.I: it is self-intersecting along the unstable horizontal ray through  $J$ , and clearly no curve can succeed in winding about such cylindrical surface without intersecting this unstable ray -- a possibility ruled out by the uniqueness theorem of the differential equations controlling the rays\*. On the other hand, rays close to the stable horizontal rays (which cut  $\Sigma$  in  $D$  and in  $D_1$  and  $D_2$ ) wind about their simple closed cylindrical surfaces parallel to the stable

---

\* Cf. e.g., Goursat & Hedrick, Vol. 2.<sup>5</sup>

horizontal ray, which is inside the cylinders. During the flow in  $\Sigma$ , their representative points revolve about the fixed representative points  $D, D_1, D_2$ . They are evidently periodic, as shown analytically in Appendix B, but with periods depending on  $k$ .

Certain special  $k$ -constant curves have been marked on Figures 14.III and 14.IV:  $K_0$  forms the lateral boundary of  $\Sigma$ ;  $K_1$  bounds the "inner core" (to be explained below); rays whose representative points are between  $K_0$  and  $K_1$  are important in bottom bounce and RAP propagation at short ranges. Rays represented by points between  $K_1$  and  $K_2$  propagate to long ranges by virtue of surface reflections (RSR propagation). Rays represented by points within  $K_2$  remain entirely within the sound channel.

In the non-laminar case, these sharp distinctions disappear, although represent an approximation.

We may note that if  $\tilde{\mathcal{S}}_3$  had been represented in the coordinates  $(r, z, \gamma)$  instead of  $(r, z, p_z)$ , all would have been similar to the above, except that the lateral boundaries are vertical planes  $\gamma = 0, \pi$ ; and  $I_2$  would have had  $p_r \delta z \wedge \delta \gamma$  for integrand: the transformation preserves "mass" corresponding with  $p_r$  as density, not area. See Appendix E.

Power flow in the surface of section: intermediate and long ranges. As shown in Section 13, the function  $p_r \tilde{f}$  of (13.11) is constant along each ray in  $\tilde{\mathcal{S}}_3$ , except for those rays which meet the bottom, when it drops to zero -- according to the assumption of total absorption. This means that the only part of  $\Sigma$  relevant to intermediate and long range propagation is what remains after the representative points of such terminating rays have been removed. Calculations based on actually

observed acoustic profiles show that most of the area of  $\Sigma$  is removed in this process: only the part corresponding with the "central core" (the acoustic channel) is involved in intermediate and long range propagation. It is to this central core of  $\Sigma$  that the rest of the developments in this section apply. Of course for ranges of one or two convergence zones, rays corresponding to the omitted part of  $\Sigma$  are the important ones, entering, as stated before, into the phenomena of the reliable acoustic path (RAP) and bottom bounce, so important in acoustic detection at these shorter ranges. The central core is bounded by  $K_1$ :  $k = n(b)^*$ .

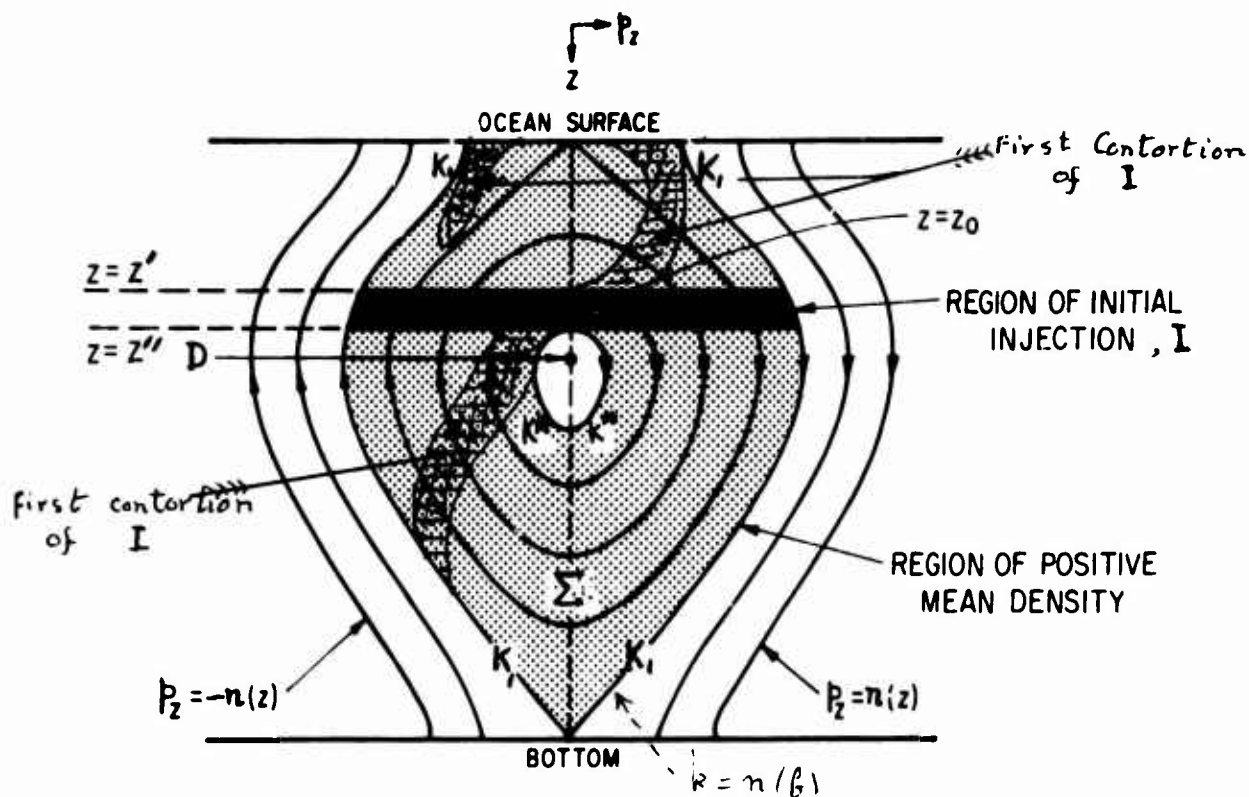
We incorporate the power flow density  $\tilde{f}$  in the expression

$$(14.4) \quad g = g(r, z, p_z) = p_r r \tilde{f}(r, z, p_z)$$

which, as we have said, is constant along each ray cutting the central core of  $\Sigma$ : during the flow induced in it by the rays, each representative point carries its individual value of  $g$  along with it; and each "level curve"  $g = \text{const.}$  in general deforms and moves in  $\Sigma$  as the flow progresses. Initially, i.e., at a range  $r = r_0$  close to the emitter A, the values of  $g$  are simple to describe: if the depth interval  $(z', z'')$  spans the height of the emitter together with its neighboring waters contributing radiation into the sound channel, then since the range of directions  $\gamma$  pointing into it correspond to the values of  $p_z$  at the boundaries of the central core — i.e., to  $\pm p_z(z)$ , where  $p_z(z) = \sqrt{n^2(z) - n^2(b)}$ , then the part of  $\Sigma(r_0)$  over which  $g$  is not zero is the thin horizontal slice on which

$$(14.5) \quad z' < z < z'', \quad -p_z(z) < p_z < p_z(z).$$

\* For variable bottom  $b = b(r)$ ,  $b$  is replaced by its minimum. CF. however, the sea-mounts examined later.



**FIGURE 14.V - Evolution of Mean Density with One Duct**

This shall be called the emitter's injection region, I. Our assumptions regarding the approximately uniform radiation in directions entering the sound channel and the azimuthal symmetry have as consequence the approximate constancy of  $g = g_0$ , except near its horizontal edges at  $z'$  and  $z''$ , where it falls to zero quite abruptly but continuously. It is necessary to picture the evolution of this injection region  $I \rightarrow I(r)$  in  $\Sigma(r)$  during the course of the flow; i.e., the increase of range  $r$ .

The rays through the injection region wind about the axis of the duct as suggested by Figure 14.I in the laminar case; and similarly, but less evenly, in the more general one. In the former case each ray

is periodic, as shown in Appendix B; but with a period  $R$  that depends on the Snell constant:  $R = R(k)$ ; therefore as  $r$  increases from  $r_0$  to  $r_0 + R$  (i.e., by the period of a particular ray of the injection region) the points of this region will not in general return to their former positions -- only those will whose period happens to be  $R$ . Hence the image of the injection region in  $\Sigma(r_0 + R)$  will be the result of advancing points on each curve  $K$ :  $p_r = k$  a different amount, depending on  $k$ . This leads to the contorted image indicated schematically in Figure 14 V. As  $r$  is further increased, this contortion increases, so that the injection region tends to invade the whole part of  $\Sigma$  between the extreme curves  $K_1$  and  $K^*$  corresponding to Snell constant extremes. Since the rays in acoustic channels have been shown, under a wide range of actual conditions, to have periods varying by 30% or more depending on their Snell constant (as given by their initial directions), it is clear that, as the range reaches some three convergence zones, the image of the injection region is twisted all around in the sub-region of  $\Sigma(r)$  to which it is confined. This subregion is the area between the curves  $K_1$ ,  $K^*$ , and the loci of (14.3) for the extreme values of  $k$  in the injection region, shown graphically in Figure 14.V. The calculation of these extreme values of  $k$  is done by substituting the limiting values  $(z', p_z')$  and  $(z', p_z = 0)$  into (14.3), etc. Since the image of the injection region contorts increasingly as  $r$  increases still further, the picture that develops is one of its penetrating throughout the whole area of  $\Sigma(r)$  between  $K_1$  and  $K^*$ , so that every point is either in the image or close to points in the image. We have stated the matter as it applies to the laminar case. It shows why it is natural to take the lower limit of the "intermediate ranges" as

of the order of three convergence zones. In the non-laminar case the picture on  $\Sigma(r)$  is at least as irregular, but the images of the injection region are not precisely between curves such as  $K_1$  and  $K^*$ : see Section 16.

We confine the rest of this section to the laminar case, deriving consequences from the considerations of the last paragraph: the dependence of period  $R = R(k)$  on Snell constant  $k$ , so that points on different Snell curves are advanced by different amounts, the cumulative effects contorting  $I(r)$  more and more so that it penetrates throughout the central core of  $\Sigma(r)$  as  $r$  increases. [Secs. 15, 16 extend the results to  $W \neq 0$ .]

Let us fix a small region  $\Delta\Sigma$  on the central core of  $\Sigma(r)$ ; i.e., a region which does not vary with  $r$ . As the flow on  $\Sigma(r)$  progresses, the contorted  $I(r)$  will intersect  $\Delta\Sigma$  in a region  $\Delta I(r) = I(r) \cap \Delta\Sigma$  whose shape and area will vary with  $r$ . We will assume that the area of  $\Delta I(r)$  approaches a limit as  $r$  increases, and, in fact, is close to this limit after relatively few convergence zones. A similar assumption was made by J.W. Gibbs<sup>6</sup> and by H. Poincaré<sup>\*</sup>, who based it on the intuitive analogy of the visible behavior of liquids on stirring or of smoke in a steady circulatory motion of air, and applied it to statistical mechanics. Only relatively recently have theorems been established which place such assumptions on a rigorous basis; they belong to ergodic theory; an outline of these questions in their bearing on the present situation will be found in Appendix G. \*\*

Taking a fixed point  $(z, p_z)$  in our fixed region  $\Delta\Sigma$ , let us consider the behavior of our  $\lim_{r \rightarrow \infty} \Delta I(r)$  as  $\Delta\Sigma$  shrinks down to it. It seems clear that for small  $\Delta\Sigma$ ,  $\lim_{r \rightarrow \infty} \Delta I(r)$  is proportional to the area

\*In his lectures on probability.

\*\*cf. second, third, fourth and fifth references.<sup>6</sup>

of  $\Delta\Sigma$ , the coefficient being a function of  $(z, p_z)$ ; more precisely, it would seem that there is a density  $h$  defined by

$$(14.6) \quad h(z, p_z) = \lim_{\Delta\Sigma \rightarrow 0} \lim_{r \rightarrow \infty} \frac{\Delta I(r)}{\Delta\Sigma}$$

This formula expresses the density  $h$  as a double limit, and, as J.W. Gibbs noted, would not exist if the order of taking limits in (14.6) were inverted (i.e., first  $\Delta\Sigma \rightarrow 0$  then  $r \rightarrow \infty$ ). The result expressed in (14.6) can also be justified by the ergodic theory -- given certain natural assumptions.

It is easy to see that  $h(z, p_z)$  is constant along each ray in  $\Sigma_3$ ; for let  $(r, z, p_z)$  and  $(r', z', p'_z)$  be two points on the same ray, and let the small region  $\Delta\Sigma$  be traced in  $\Sigma(r)$  containing  $(z, p_z)$  and  $\Delta\Sigma'$  be its image in  $\Sigma(r')$ : it will contain  $(z', p'_z)$ , and have the same area as  $\Delta\Sigma$  -- by the area-preserving property of our transformation on  $\Sigma(r)$ . By the same property, the area of  $\Delta I(r)$  is equal to that of its image,  $\Delta I(r')$ . Thus the fraction  $\Delta I(r)/\Delta\Sigma$  in (14.6) is equal to the corresponding fraction  $\Delta I(r')/\Delta\Sigma'$  for the image  $(r', z', p'_z)$  of  $(r, z, p_z)$ . Hence in the limit  $h(z, p_z) = h(z', p'_z)$ , as was to be shown.

As the point  $(z, p_z)$  moves in  $\Sigma(r)$  with increasing  $r$ , its path is, as we have seen, one of the curves of given Snell constant  $k$ . Since it carries the value of  $h(z, p_z)$  along with it, this function is constant along every such curve in Figure 14.V, and on each connected piece of a Snell curve in Figure 14.IV. This means that with a single duct, when the Snell constant  $k$  is given, the value of  $h(z, p_z)$  is determined -- in other words, the latter is a function of the former:

$$h(z, p_z) = F(k).$$

In the case of two ducts, this function could be double-valued, having one value on one connected piece (or half of the figure eight), and a different value on another connected piece of the curve for a given  $k$ : instances of this occur in certain cases of shadow zones. Finally, there is the important possibility that the above function be constant -- at least throughout certain of its regions bounded by Snell curves. This would imply that  $F(k)$  is discontinuous, e.g., a step function. It is in this way -- and with  $h$  dropping to zero in some sub-regions of  $\Sigma$  -- that shadow zones manifest themselves.

The general expression of  $F(k)$  in terms of the given quantities and its evaluation in special cases are reasonably simple problems, and will be solved below. But first we wish to show that once  $F(k)$  is found, the acoustic intensity at any point  $(r, z)$  of the ocean and the transmission loss (TL) to that point can be calculated.

The method is based on the assumption discussed above, namely, that the quantity  $g(r, z, p_z)$  defined in (14.4) -- and which was shown in (13.12) to be constant along each individual ray -- is also constant all over the initial injection region I. Let its value be denoted by  $q$ .

First, we can express the acoustic intensity over I (i.e., near the source) in terms of  $q$  -- and vice versa. For this purpose we apply (13.10) to this region,  $C$  taken as the vertical segment  $z' < z < z''$  of (14.5), and the domain  $\Gamma$  of integration being the angular interval

$$(14.7) \quad \Gamma: \gamma_0 < \gamma < \pi - \gamma_0, \quad \gamma_0 = \sin^{-1}[n(b)/n(z)]$$



which corresponds to the second inequality in (14.5). The angle of the normal  $\gamma' = \pi/2$ , so that  $\cos \Theta = \sin \gamma$ . Since, further,

$r_0 p_r^0 \tilde{f}(r_0, z_0, p_z^0) = q$ , equation (13.10) gives—with the aid of (13.6)—the result

$$P_2 = q \int_{z'}^{z''} n(z) dz \int_{\gamma_0}^{\pi-\gamma_0} \sin \gamma d\gamma = 2q \int_{z'}^{z''} n(z) \cos \gamma_0 dz$$

Applying the law of the mean to the last integral, it is expressed exactly as  $2q(z''-z') p_z(z_0)$ , where  $z_0$  is suitably chosen in the interval  $(z', z'')$ —and with acceptable approximation if it is taken at the mid-point, as we shall do henceforth. The resulting  $P_2$  is the power flux across  $I$  per unit azimuth ( $\phi$ ). To find the power flux per unit azimuthal arc ( $s_0 = r_0 \phi$ ) we must divide the expression by  $r_0$ . Finally, to find the power flux per unit area, we divide this further result by  $z''-z'$ . Now the acoustic intensity is the power flux per unit area across an element of area for which this has a maximum value. In the present case this coincides with an element of our vertical cylindrical surface, generated by revolving the depth interval  $(z', z'')$  about the  $z$ -axis (taken through the emitter). This follows from the calculations in the above evaluation of  $P_2$ . Hence

$$(14.8) \quad \text{acoustic intensity on } I = 2q p_z(z_0)/r_0.$$

After these preliminaries, we evaluate the corresponding quantity at an arbitrarily given point  $(r, z, p_z)$  in the central core and at a long enough range for (14.6) to give an acceptable approximation. The whole matter turns on the replacement, in (13.10), of the density  $\tilde{f}$  by a sort of average,  $\bar{f}$ , corresponding with the limiting processes underlying (14.6). The reasoning is as follows:

Let  $\Delta\Sigma$  be a small fixed region in  $\Sigma(r)$ . At those points of  $\Delta\Sigma$  through which rays pass that cut I—i.e., at points of  $\Delta I(r)$ —the value of  $\tilde{f}$  is given by the equation derived from (14.4) and the assumption  $g=q$ : therefore  $\tilde{f}=q/rp_r$ ; at all other points of  $\Delta\Sigma$ ,  $\tilde{f}=0$ . Therefore the total power flux through  $\Delta\Sigma$  is this  $\tilde{f}$  multiplied by the area  $\Delta I(r)$ ; and the power flux per unit  $zp_z$ -area is found by dividing this product by area  $\Delta\Sigma$ , giving  $\tilde{f} \Delta I(r)/\Delta\Sigma$ . The result of replacing this quotient of areas by its limit, (14.6), i.e., by  $h(z, p_z)=F(k)$ , is the mean flux density, and we have

$$(14.9) \quad \tilde{f}(r, z, p_z) = \frac{q}{rp_r} F(k).$$

We now insert this into equation (13.10) in place of  $\tilde{f}$ , and integrate over an angular interval  $\Gamma_1$ , defined as in (14.7), but now using for  $z$  the depth at the receiver, which lies within its depth interval  $(\zeta', \zeta'')$ . We obtain

$$(14.10) \quad P_2 = q \int_C d\sigma \int_{\Gamma_1} n F(k) \cos \theta \, d\gamma$$

To express this power flux per unit azimuth as a flux per unit azimuthal arc, we divide by  $r$ , as before. Now select for  $C$  an element of arc,  $\Delta C$ , so that the  $\sigma$ -integration reduces to multiplication by the latter. When the result is divided by  $\Delta C$ , the following expression is obtained for the power flux per unit area (in physical space):

$$(14.10)' \quad \begin{aligned} \text{pr. flux unit area} &= (q/r) \int_{\Gamma_1} n \tilde{F}(k) \cos (\gamma - \gamma') \, d\gamma \\ &= R(r, z) \sin \gamma' + Z(r, z) \cos \gamma' \end{aligned}$$

where  $R$  and  $Z$  are the radial and vertical components of this flux vector (cf. the  $U, V, W$ , in (11.5)). We will show that  $Z = 0$ , so that the acoustic intensity at the point  $(r, z)$  will be simply  $R$ , given below.

In the integration over  $\Gamma_1$ , the angle  $\gamma$  goes from a value  $\gamma_1$  for which  $k = n(z) \sin \gamma_1 = n(b)$ , so that  $\sin \gamma_1 = n(b)/n(z)$ ; through the value  $\gamma = \pi/2$  ( $k = n(z)$ ); and thence to  $\pi - \gamma_1$ . This can be seen from Figures 13.III and IV, etc. In these two symmetrical sub-intervals of  $\Gamma_1$ ,  $\sin \gamma$  and  $k$  run through the same (positive) values; whereas  $\cos \gamma$  and  $p_z$  go through equal and opposite values. The explicit expression of  $Z(r, z)$  obtained from (14.10) shows that the integrand is  $n(z)F(k)\cos\gamma = p_z F(k)$ ; and this, being reversed in sign in the second sub-interval as compared with the first, leads to a zero integral over  $\Gamma_1$ , proving our statement that  $Z = 0$ .

For  $R$  we have, applying (14.8), the expression

$$(14.11) \quad R = \frac{q}{r} \int_{\gamma_1}^{\pi - \gamma_1} F(n \sin \gamma) n \sin \gamma \, d\gamma$$

From the symmetry of the integrand in the two sub-intervals of  $\Gamma_1$ , it is seen that this integral may be written as twice the integral over the first sub-interval  $(\gamma_0, \pi/2)$ , which corresponds to  $k$  running from  $n(b)$  to  $n(z)$ . Hence, writing

$$n \sin \gamma \, d\gamma = -d(n \cos \gamma) = -dp_z = kdp/p_z$$

$$(14.12) \quad R = \frac{2q}{r} \int_{n(b)}^{n(z)} F(k) \frac{kdk}{p_z(z,k)} = \text{acoustic intensity at } (r,z).$$

Since the vertical component  $Z$  is zero, the maximum power flux per unit area is in the radial direction; hence it is the acoustic intensity, as stated.

The transmission loss is the ratio of the acoustic intensity at the given point  $(r,z)$  to that at the source e.g., on  $I$ . From (14.8) and (14.12) it is

$$(14.13) \quad TL = \frac{r_o}{r} \frac{1}{p_z(z_o)} \int_{n(b)}^{n(z)} F(k) \frac{kdk}{p_z(z,k)}.$$

[Note that "TL" is most frequently measured in decibels, so that "TL" =  $10 \log_{10} TL$ , given in (14.13)].

It remains to find the function  $F(k)$ . For this purpose, we first observe that (14.6) expresses the function  $h(z, p_z)$  as a density over the plane  $\Sigma$ , which, when integrated over any fixed sub-region  $\Lambda$  of  $\Sigma$ , gives the limiting area of intersection of  $I(r)$  with  $\Sigma$ :

$$(14.14) \quad \int_{\Lambda} \int h(z, p_z) dz dp_z = \lim_{r \rightarrow \infty} \text{area of } \Lambda \cap I(r)$$

[proof: subdivide  $\Lambda$  into a large number of small pieces  $\Delta\epsilon$ , add corresponding products  $h(z, p_z)\Delta\epsilon$ , and pass to the limit].

We shall take  $\Lambda = \Lambda(k)$ , the locus of points of  $\Sigma$  for which  $p_r \geq k$ . Since the maximum value of  $k$  is  $n(\bar{z})$ , where  $\bar{z}$  is the depth at which  $n(z)$  is maximum (minimum sound speed  $c$ ), we have  $k \leq n(\bar{z})$ . When  $k = n(\bar{z})$ ,  $\Lambda(k)$  reduces to the point  $(\bar{z}, 0)$ , corresponding to the stable horizontal ray. As  $k$  decreases from  $n(\bar{z})$  to its minimum value  $n(b)$  for the central core,  $\Lambda(k)$  grows from this point to the whole central core, bounded by  $K_1$  of Figures 14.III, IV, V. The area of  $\Lambda(k)$  — also to be denoted by  $\Lambda(k)$  — is, as was seen earlier, the contour integral of  $p_z \delta z$  about its boundary ( $p_r = k$ ) in the negative sense with respect to the enclosed region  $\Lambda(k)$ .

Consider the marginal region  $\Delta\Lambda = \Lambda(k - \Delta k) - \Lambda(k)$ , i.e., the locus of points for which  $k - \Delta k \leq p_r < k$ : it is bounded by Snell curves of the types shown in the above figures. In the simplest case of a single relative maximum of  $n(z)$  (III, V),  $\Delta\Lambda$  is either a ring, or, in the region RSR propagation, the lower part of such a ring, cut off by the ocean surface  $z = a$ ; but more complicated situations may occur. In all cases, however, the area of  $\Delta\Lambda$  is given by contour integration of  $p_z \delta z$  about its whole boundary.

Since the "Snell curves" on  $\Sigma$  are invariant under the transformations in it (produced by varying  $r$ ), both  $\Lambda(k)$  and  $\Delta\Lambda$  are invariant regions in this "flow". Therefore their figure of intersection with  $I(r)$ , while contorting and changing with  $r$ , always remains in  $\Delta\Lambda$  and has an unchanging area. Therefore the "limit" sign in (14.14) when  $\Lambda$  is taken as this region  $\Delta\Lambda$ , may be dropped: i.e., the integral is equal to the constant value of the area of  $\Delta\Lambda \cap I(r)$ . Taking, in particular,  $r=r_0$ , the value is the area of the small figure cut off by lines of depths  $z'$  and  $z''$  from the thin marginal  $\Delta\Lambda$ ; cf. the curvilinear quadrilaterals in Figure 14.V. Whatever its shape, the area is always given by contour integration of  $p_z \delta z$ .

On the other hand, since our basic (ergodic) assumption is that  $h(z, p_z) = F(k)$ , the integral on the left in (14.12) is easily found, when  $\Lambda = \Delta\Lambda$ , to be the area of  $\Delta\Lambda$  times  $F(\bar{k})$  where  $\bar{k}$  is between  $k - \Delta k$  and  $k$ . Equating this product to the expression of the last paragraph, we obtain

$$(14.15) \quad F(k) = \lim_{\Delta k \rightarrow 0} \frac{\text{area of region } \{z' < z < z'', k - \Delta k < p_r < k\}}{\text{area of region } \{k - \Delta k < p_r < k\}}$$

This limit may not exist for certain exception values of  $k$ . If  $I$  and  $\Delta\Sigma$  do not intersect, it is zero (shadow zone); if they do, it is in general positive and finite; but at possible discontinuities at boundaries of shadow zones, no limit may be approached.

We shall evaluate  $F(k)$  at a non-exceptional point, and in the simple illustrative case of Figure 14.V. The denominator in (14.15), being the difference in the areas,  $\Lambda(k - \Delta k) - \Lambda(k)$ , is to quantities of higher order,  $-\Lambda'(k)\Delta k$ . Now evidently

$$\Lambda(k) = 2 \int_{z_1}^{z_2} |p_z| dz = 2 \int_{z_1}^{z_2} \sqrt{n^2 - k^2} dz,$$

where  $z_1$  and  $z_2$  are the intersections with the  $z$ -axis of the curve  $p_r = k$  (i.e.,  $n(z_1) = k$ ). When differentiated with respect to  $k$ , three terms are obtained, two contributed by the limits of integration (which are functions of  $k$ ) — but these are both zero. The third term is given by differentiation under the integral sign (Leibnitz' rule). [its integrand becomes infinite at the two limits of integration, but, in the general case when  $n'(z_1) \neq 0$ , being a uniformly convergent improper integral, this formal process is valid]. Thus, finally, dropping higher order terms in  $\Delta k$ ,

$$(14.16) \quad \Delta\Lambda = -\Lambda'(k)\Delta k = \Delta k 2 \int_{z_1}^{z_2} \frac{k dz}{\sqrt{n^2 - k^2}} = R(k) \Delta k$$

where  $R(k)$  is the period of the ray; see Appendix B, in particular Figure BI and equation (B.6).

Similarly, the area in the numerator is given by an integral like that for  $\Delta\Lambda(k)$ , but with  $z'$  and  $z''$  as its limits of integration. To quantities of higher order in  $\Delta k$  it is an integral like (14.16), but again with the above change in limits of integration. Therefore we

may cancel the  $\Delta k$  from numerator and denominator in (14.15), whereupon we see that  $F(k)$  is the ratio of the two integrals in question. Since the depth interval of the emitting body,  $z'' - z'$ , is small, we write the numerator in the form

$$(z'' - z') \frac{2k}{\sqrt{n^2(z_0) - k^2}} \quad (z' < z_0 < z'')$$

This expression, exact for a suitably chosen  $z_0$  (by the law of the mean), is correct to terms of higher order in  $z'' - z'$  for any  $z_0$  in the interval. Thus we have

$$(14.17) \quad F(k) = (z'' - z') \frac{2k}{\sqrt{n^2(z_0) - k^2}} \cdot \frac{1}{R(k)},$$

with  $R(k)$  given by (14.16).

It is useful to express this result in terms of the "aperture" of the region of initial injection I: Taking as before  $z_0 = (z'' - z')/2$ , the point A on the axis of our cylindrical coordinates with  $z = z_0$ ,  $r=0$ , may be described as a central reference point of the emitter. The angle subtended at A by the two points A': ( $r=r_0$ ,  $z=z'$ ) and A'': ( $r=r_0$ ,  $z=z''$ ) may be called the "aperture" of I; it is determined by

$$\angle A'AA'' = 2a, \quad \tan a = (z'' - z')/2r_0.$$

Hence (14.17) becomes

$$(14.18) \quad F(k) = \frac{4k}{p_z(z_0, k)} \cdot r_0 \tan a \cdot \frac{1}{R(k)}$$



where  $p_z(z_0, k) = \sqrt{n^2(z_0) - k^2}$ , as usual.

With this expression for  $F(k)$  inserted into (14.12) and (14.13), explicit formulas for the acoustic intensity and the transmission loss are obtained. We write the latter in the form

$$(14.19) \quad TL = \frac{\tan a}{p_z(z_0, k)} \frac{r_0^2}{r} \int_{n(b)}^{n(z)} \frac{4k^2 dk}{R(k) p_z(z_0, k) p_z(z, k)}$$

This can be simplified by making the following physically reasonable (and usual) assumption, making the choice of the vertical segment  $(z', z'')$  more precise:

The aperture is equal to the angle  $(2\alpha_0)$  between the two critical rays.

Then  $p_z(z_0, k) = n(z_0) \cos \gamma_0 = n(z_0) \sin \alpha_0 = n(z_0) \sin a$  so that, in (14.19),  $(\tan a)/p_z(z_0) = c(r_0) \sec a$ . Since for the angles as close to zero as occur here,  $\sec a = 1 + [a^2] \approx 1$ , this reduces to  $c(z_0)$ , and we obtain the formula (with the standard convention  $r_0 = 1$ ).

$$(14.20) \quad TL = \frac{c(z_0)}{r} \int_{n(b)}^{n(z)} \frac{4k^2 dk}{R(k) p_z(z_0, k) p_z(z, k)}$$

This formula coincides with the laminar case of the expression (25.1) for the transmission loss in the correlative study of the present series: "Computation of Long Range Propagation Losses in a Duct" (June 20, 1973). The following line-up of symbols is observed,

the first line being those used in the present study, the second, in the one cited:

$r$	$z$	$z_0$	$p_r, k$	$\Lambda(k)$	$R(k)$	$n(z_0)$
$x$	$z, z_r$	$z_s$	$p$	$J$	$\lambda$	$1/c (0, z_s)$

The fact that such apparently different methods as the present one and that of the memorandum cited give the same results (in the laminar case, at least) may deserve some comments--beyond the general truism that two correct methods should give the same result. For not only are the two methods based on approximations, but on what may at first appear to be different models: the present study starts with the explicit assumption that in the radiation field infinitely many rays--one in each direction--pass through each point. The reference starts with individual rays, but in its averaging process, actually translates them horizontally, so that the result is to have, again, infinitely many rays through each point. Finally, the mode of averaging in the present paper is over the surface of section--or the rings into which it is subdivided by the Snell constant curves, whereas the method in the reference is to average by horizontal translation in space. The fact that these two methods give the same result is a corollary of the ergodic theory.

## 15. Continuation: The Statistically Laminar Case

This section continues the study of rays in a vertical plane with azimuthal symmetry, treated in the last two sections, and uses the model and the four assumptions of Section 13,<sup>\*</sup> together with the graphical representation of the phase space  $\mathcal{S}_3$  and the surface of section  $\Sigma(r)$  of Section 14. Now, however, the sound speed and refractive index  $1/c = n = n(r, z)$  will actually vary with  $r$  as well as with  $z$ . This has the two following consequences: First, the quantity  $p_r$  is no longer a first integral of the ray differential equations--its value changes along each ray (Snell's law in the large,  $p_r = k$ , is invalid). Secondly, the transformations on the surface of section  $\Sigma(r)$ , induced by the rays and produced by changing  $r$ , and which we have described as a "flow" in  $\Sigma$ , is no longer a steady flow: if when  $r$  is increased to  $r'$  the point  $(z, p_z)$  goes into the point  $(z', p'_z)$ , it will in general not be true that when  $r_1$  is increased by the same amount  $(r' - r)$  to  $r'_1 = r_1 + (r' - r)$ , the point  $(z, p_z)$  will go into  $(z', p'_z)$ . In mathematical language, our one-parameter family of transformations on  $\Sigma$  no longer has the group property.<sup>†</sup> The flow is, nevertheless, incompressible since this area-preserving property results from the integral invariants of any Hamiltonian system.

From the first consequence ( $p_r$  not constant along a ray) it follows that the rays in  $\mathcal{S}_3$  are not confined to cylindrical surfaces as in Figure 14.I, and the flow in  $\Sigma$  is not along fixed curves (lines of steady flow, the Snell curves of Figures 14.III, IV, V). And of course our ray differential equations cannot be solved by quadratures, as in Appendix B.

<sup>\*</sup> In Assumption 2,  $W$  is a random perturbation and  $\bar{W} = 0$ .

<sup>†</sup> To be regained in Section 16 by a deformation when  $W$  is non-random but small.

From the second consequence (the non-steady nature of the flow in  $\Sigma$ --the absence of the group property) a still more radical change is produced in the methods used in the laminar case: the ergodic theory does not apply. Therefore, while we may still think intuitively of the mixing of flowing regions on  $\Sigma$ , our earlier basis for such conclusions as those embodied in (14.6) and its consequences, no longer exists. On the other hand--and this will come as no surprise to students of statistical phenomena in physics--those same environmental complexities, that have deprived us of the earlier methods, actually give simpler results, since the random departures from laminarity tend to even things out at the larger ranges of concern.

First let us see how the variation of profile with range, concerning which Assumption 2 of Section 13 has been postulated ( $\bar{V} = U = U(z)$ ), yields information concerning the first order perturbations in the rays when  $V = U(z) + W(r, z)$  replaces  $\bar{V} = U(z)$  in the canonical equations. For this purpose we apply the model and the statistical methods to the perturbations given in Section 12. The reasoning leading to equations (12.2)-(12.7) gives, in the case of cylindrical coordinates, where  $U$  is independent of  $r$ ,

$$dp_r/d\tau = -\partial U/\partial r - \partial W/\partial r = -\partial W/\partial r,$$

and  $\bar{W} = 0$  (to terms of first order). Hence  $d\bar{p}_r/d\tau = 0$ , and  $\bar{p}_r = p_r^0$ , its initial value, which we may denote by  $k$ . On the other hand,

$$dp_r/d\tau = \frac{dr}{d\tau} (dp_r/dr) = p_r (dp_r/dr) = \frac{1}{2} \frac{d}{dr} p_r^2.$$

Taking population means as above, we find  $d \overline{p_r^2}/dr=0$ . Thus  $\overline{p_r^2}$  is constant along each ray. Hence the standard deviation  $\sigma(p_r)$ , which equals the square root of  $\overline{p_r^2} - (\overline{p_r})^2$ , is constant along each ray, and since its initial value is zero (all perturbations being initially zero, as noted in Section 12), the perturbation of  $p_r$  is a random quantity of zero standard deviation: in other words, a constant. This, of course, is based on the approximation of first order terms in the calculations. The fact that when second order terms are included, as in the calculation of second moments by formulas such as (12.11) and (12.12), the standard deviations in the random perturbations can be expressed as integrals of small but positive integrands, shows that the first order conclusion is false at a higher order of approximation. In fact, as we have noted at the close of Section 12, a rough but reasonable estimate would be that  $\sigma(p_r)$  increases as the square root of the range. A further application of the first order approximation to the other coordinates and moments shows that, since the term containing  $r$  has a zero population mean ( $\overline{W}=0$ ), the steadiness of the flow in  $\Sigma$  (the group property) is valid for the mean positions as it was in the laminar case.

These considerations have two important practical consequences:

First, the methods used in the laminar case continue to give approximately valid quantitative results over long enough ranges for a random departure from laminarity to set in.

Second, beyond the ranges within which the first order statistical calculations give acceptably accurate results, it is the higher order statistical quantities in the profiles--more explicitly, the variances

and covariances in the ensemble  $\{V\}$  --that give the measure of the necessary corrections. This has been indicated in Section 12. Such ranges will be termed "long ranges" to distinguish them from the "intermediate ranges", at which quasi-laminarity and ergodic mixing are acceptable, and the "short ranges" at which laminarity and deterministic calculations based on individual rays give valid methods. The import of our second conclusion for long ranges is that since the second order statistical quantities of the acoustic medium are not only unknown, but are shown to be subject to random turbulence,<sup>2</sup> their values are less stable and permanent than the first order quantities (e.g., mean sound speeds at given positions and seasons)—the appropriate methods of calculating their effects are stochastic ones.

The basis of such methods, as we shall apply them, are intuitively simple and, in their mathematical form, long familiar in the statistical mechanics of systems in equilibrium. The intuitive picture to which we appeal is that of card shuffling, when no cards are lost or added to the pack; or of stirring of a fluid, provided no matter is lost or gained. The result of such shuffling or stirring is a uniform spreading out of any initially identified set of cards throughout the pack or of a stained portion of the fluid throughout its mass. Our acoustical application is to the surface of section  $\Sigma$ , the stained portion being the region of initial injection  $I$ , and the conservational property the preservation of areas (the Hamiltonian integral invariant) and the conservation of

power flow. The result is a statistically uniform dispersion of I throughout the parts of  $\Sigma$  corresponding to the sound channel. This leads to a drastic simplification of the formulas replacing (14.13) and (14.19), giving essentially the same result as would be obtained by replacing  $F(k)$  by a constant--moreover, a constant easy to calculate.

There is one complication in the application just mentioned: In contrast to the laminar case, in which the sound channel corresponded to the central core of  $\Sigma$ , bounded by the Snell curve of rays of critical angle, in the present case such a central core can only be defined statistically, in terms of means (first order statistics). But then the variances will produce a random leakage or diffusion out of this central core: in any given range interval, some rays will cross the core's boundary and may--possibly after a surface reflection--hit the bottom and be absorbed before they regain the core. This produces a loss of power which, in our present model, is equivalent to applying a factor slightly less than unity to each range interval  $\Delta r$ , this factor being the same for any interval of the same length. The cumulative effect is to multiply the power, that would otherwise be transmitted, by the exponential factor  $e^{-\alpha r}$ , where  $\alpha$  is the loss coefficient. In the present state of our science, it is not possible to distinguish the exponential decay due to this leakage out of the duct from the possible effects of absorption and scattering, both of which multiply the transmitted power by a factor of the same type. Any experimental observation leading to the measurement of such a coefficient  $\alpha$  gives only the sum of coefficients due to all such effects.

In the quantitative expression of these largely intuitive ideas, we first define a "statistical central core"  $\bar{\Sigma}$ , where, as in Section 14, the critical angle restriction is applied by the second inequality in (14.5). However, since the quantity  $n^2(z) - n^2(b)$  used there now depends on  $r$ , we replace it by its mean  $\bar{n}^2(z) - \bar{n}^2(b)$ . The latter is obtained at once from (12.1) and Assumption 2 in Section 13, which shows that  $\bar{n}^2(r, z) = -2U(z)$ , etc., independent of  $r$ . The boundary of  $\bar{\Sigma}$  will still be denoted by  $K_1$ ; and  $I(r_0)$  will be the injection band defined by (14.5): it is bounded by  $K_1$  and the two lines of depths  $z'$  and  $z''$  as in Figure 14.V.

Into  $I(r_0)$  the total power (per unit azimuth)  $P_2$  is injected as in Section 14, and this can again be pictured as the uniform staining of this region by a dye of total amount  $P_2$ . Again in the flow of the (2-dimensional) liquid in  $\bar{\Sigma}(r)$ , as  $r$  increases from  $r_0$  to large values, the dye is transported and contorted; but now there is only a steady flow in the mean approximation: we may picture the situation as a steady flow combined with a diffusion of the stain into the unstained parts of the liquid, adding to the spread of the pigmentation, whose density on  $\bar{\Sigma}(r_0)$  is  $g(r_0, z, p_2) = q$  on  $I$ ,  $= 0$  off  $I$ ; cf. (14.8).

The picture that unfolds seems intuitively clear--and will be given a mathematical basis in the remainder of this section: at long ranges the pigmentation becomes uniform over the region  $\bar{\Sigma}(r)$ . (Note that  $\bar{\Sigma}(r_0)$  and  $\bar{\Sigma}(r)$  are distant parallel regions, both congruent



to  $\bar{\Sigma}$ ). At such ranges the intensity of pigmentation approaches a limit  $\bar{g}(z, p_z)$  -- and in the following sense: First, the quantity of dye in any fixed element  $\Delta\Sigma$  of  $\bar{\Sigma}(r)$  approaches limit  $\Delta\mathcal{G}$  as  $r \rightarrow \infty$ ; second,  $\Delta\mathcal{G}/\Delta\Sigma$  approaches  $\bar{g}(z, p_z)$  as  $\Delta\Sigma \rightarrow (z, p_z)$ . [This is the same sort of double limiting process as was used in (14.6) and in the replacement of  $\bar{f}$  by  $\bar{f}$  in the proof of (14.9) and (14.10)]. Now the nature of this uniform mixing and diffusion throughout  $\bar{\Sigma}$  leads intuitively to the conclusion that  $\bar{g}(z, p_z)$  is a constant--i.e., is independent of  $(z, p_z)$ . Since the total power per unit azimuth is  $P_2$ , this constant limiting density must be  $P_2/\bar{\Sigma}$ . On the other hand, the relation between  $q$  and  $P_2$ , derived in the proof of (14.8), is  $P_2 = 2q(z''-z')p_z(z_0)$ ; we have

$$\bar{g}(z, p_z) = 2q\Delta z p_z(z_0)/\bar{\Sigma}.$$

Because of the uniformity of this density over  $\bar{\Sigma}$ , the dyed fluid (power per unit azimuth) received at a detector having a reception band  $I_1$  (defined as  $I$  was, but at the detector's depth  $z$  and depth interval  $(z', z'')$  of length  $\Delta z$ ) is the product of this area, i.e.,  $2\Delta z p_z(z)$ , by the above, namely

$$2q\Delta z p_z(z_0) \cdot 2\Delta z p_z(z)/\bar{\Sigma}$$

The power flow through  $I_1$  in the azimuthal interval  $\Delta\phi = \Delta s_\phi/r$  is the product of the above quantity by this  $\Delta\phi$ . The acoustic intensity at the receiver is the result, divided by the element of area,  $\Delta z \Delta s_\phi$ , i.e.,

$$\frac{q}{r} 2\Delta z p_z(z_0) 2p_z(z)/\bar{\Sigma}$$

Finally, to find the transmission loss we divide this by the acoustic intensity at the source, given by (14.8), obtaining:

$$TL = \frac{r_0}{r} 2\Delta z p_z(z) / \bar{\Sigma}$$

which is exactly what (14.13) would give if we set  $F(k) = \text{const.} = I_0 / \bar{\Sigma}$ .

We now give a rigorous mathematical treatment substantiating--and in fact extending--the results of the intuitive picture just developed. The basis of the mathematical attack on the problem consists in a well known theorem in analysis, the proof of which is outlined below and in Appendix G, and which states that under the application of our transformation in  $\bar{\Sigma}$ --now involving random factors--a certain logarithmic integral tends to decrease toward a minimum. Therefore after a steady state has been approximately reached (i.e., at long ranges), this integral will be at or close to its minimum.<sup>6</sup> By a second theorem in analysis, the situation when this integral is minimum is the one of uniform distribution of  $I$  through  $\bar{\Sigma}$  mentioned in our intuitive description. The logarithmic integral employed is essentially the "information" of modern communication theory, and its negative is J.W. Gibbs' "entropy". The method based on this integral is that by which Gibbs established his "canonical distribution", and--in a much more complicated setting--Boltzmann established his "H-theorem".

For brevity we shall represent points  $(z, p_z)$  on  $\bar{\Sigma}$  by single capital letters such as  $M$ , etc.; and the element  $dz dp_z$  (or  $\delta z \delta p_z$ ) of area in an integral over  $\bar{\Sigma}$  or a piece of it by the corresponding  $dM$ , etc. Finally, in integrations over the whole of  $\bar{\Sigma}$ , the symbol for this domain of integration will not be written.

Let  $M$  be a point on  $\bar{\Sigma}(r)$ . For each individual  $V$  of the population  $\{V\}$ , a 1-1 deterministic area-preserving transformation of  $\bar{\Sigma}(r)$  onto  $\bar{\Sigma}(r')$  ( $r' > r$ ) is determined, by which  $M$  is transformed into a definite point,  $M'$ , on  $\bar{\Sigma}(r')$ . For a different individual in  $\{V\}$  there will be a different image  $M'$ . Because of the "randomness" of the choice that Nature makes of  $V$  in each actual case, all that can be said of the image  $M'$  is that, given a piece of  $\Delta\Sigma$  of  $\bar{\Sigma}(r')$  there is a definite probability  $\text{prob}(M, r; \Delta\Sigma, r')$  that the image  $M'$  on  $\bar{\Sigma}(r')$  of the given point  $M$  on  $\bar{\Sigma}(r)$  shall belong to  $\Delta\Sigma$ .

While all the results obtained in the sequel would be valid if this probability--which is an "additive set function" of  $\Delta\Sigma$ --were used directly, it is physically justifiable, and contributes to the familiarity of treatment, if we assume the existence of a probability density

$$Q(M, r; M', r') = \lim_{\Delta\Sigma \rightarrow M'} \text{prob}(M, r; \Delta\Sigma, r') / \Delta\Sigma,$$

so that, for any subregion  $\Sigma^*$  of  $\bar{\Sigma}(r')$

$$(15.1) \quad \text{prob}(M, r; \Sigma^*, r') = \iint_{\Sigma^*} Q(M, r; M', r') dM'.$$

Clearly if what is given is not a precise position  $M$  on  $\bar{\Sigma}(r)$  but a probability distribution of such positions, of density  $P(M)$ , where, for any part  $\Sigma^*$  of  $\bar{\Sigma}$ ,

$$\iint_{\Sigma^*} P(M) dM = \text{prob}(M \text{ is on } \Sigma^* = \Sigma^*(r)),$$

the above formulas show us how to calculate the new probability distribu-

tion  $P'(M')$  of images on  $\bar{\Sigma}(r')$ : by compound and total probability (and the usual subdivision of regions, combination of possibilities, limit taking, etc) we obtain

$$(15.2) \quad P'(M') = \iint P(M) Q(M, r; M', r') dM$$

This is the familiar expression for changing probability densities in stochastic processes. When no confusion results, we shall drop the letters  $r$  and  $r'$ , and write the factor in the integrand as  $Q(M, M')$ ; it is called the "kernel" of the transformation; it is obviously non-negative.

Since the image of any point  $M$  on  $\bar{\Sigma}(r)$  is somewhere on  $\bar{\Sigma}(r')$ --by our earlier assumption--it follows by total probability that the integral of our kernel with respect to the second variable is unity:

$$(15.3) \quad \iint Q(M, M') dM' = 1.$$

On the other hand, there is no reason adduced up to this point that would permit us to assume that its integral with respect to the first variable  $M$  should have any particular value, or indeed be independent of  $r$ . This is because we have not yet made use of the area-preserving property of the transformation (for each choice of  $V$ ) on the surface of section. We now prove the following:

Theorem 1. For the present kernel  $Q(M, M')$ ,

$$(15.4) \quad \iint Q(M, M') dM = 1$$

The basis of the proof is the following:

Lemma. Formula (15.2) transforms a uniform distribution on  $\bar{\Sigma}(r)$ :  $P(M) = 1/\text{area } \bar{\Sigma} = 1/\bar{\Sigma}$  into the same uniform distribution on  $\bar{\Sigma}(r')$ :  $P'(M') = 1/\bar{\Sigma}$

Clearly if (15.2) is applied with

$$P(M) = P'(M') = 1/\bar{\Sigma}$$

the equation (15.4) follows, by cancelling this reciprocal area. To establish the lemma it is evidently sufficient to prove that each deterministic transformation corresponding to each individual choice of  $V$  from the ensemble  $\{V\}$  leaves the uniform distribution invariant. Let  $\bar{\Sigma}$  be any piece of  $\bar{\Sigma}(r)$  and  $\bar{\Sigma}'$  its image in  $\bar{\Sigma}(r')$ . Since a point  $M$  will be in  $\bar{\Sigma}$  if and only if its image  $M'$  is in  $\bar{\Sigma}'$ , the (unconditional) probabilities of each of these events are equal. Since the areas of the regions  $\bar{\Sigma}$  and  $\bar{\Sigma}'$  are also equal (by the Hamiltonian property) we have

$$\frac{\text{prob}(M \text{ in } \bar{\Sigma})}{\text{area } \bar{\Sigma}} = \frac{\text{prob}(M' \text{ in } \bar{\Sigma}')}{\text{area } \bar{\Sigma}'}$$

From this, by taking limits as  $\bar{\Sigma}$  and  $\bar{\Sigma}'$  approach  $M$  and  $M'$  respectively, we get the equality of probability densities at  $M$  and its image  $M'$ . In its general form this result is useless, since  $M'$  depends on the randomly chosen  $V$ . But if applied to the case in which the density

approached by the quotient on the left is uniform, it shows that the one approached on the right is also uniform--and this for every choice of  $V$ --which proves our lemma.

We now have the classical theorem of information theory, based on the "convexity" properties of the function  $u \log u$  (defined by taking the natural logarithm when  $u$  is positive, and as its limit, zero, as  $u \rightarrow 0$  through positive values). We write the non-linear functional

$$(15.5) \quad G[P(M)] = \iint P(M) \log P(M) \, dM$$

and have the theorem:

Theorem 2. For every transformation of type (15.2) in which the kernel  $Q(M, M')$  satisfies (15.4),

$$(15.6) \quad G[P'(M')] \leq G[P(M)],$$

the equality occurring only when  $P(M)$  minimizes  $G[P(M)]$ , or else when (15.2) is "trivial" (deterministic).

The proof, in brief outline, consists in observing that, in virtue of (15.4), transformation (15.2) replaces  $P(M)$  by a weighted mean  $P'(M; r) = \sum P(M) Q(M, M')$ , where the kernel  $Q(M, M')$  is the "weighting factor"--actually, a set of such factors, depending on  $r, r', M'$ . By the convexity of  $u \log u$  (graph concave up for all  $u \geq 0$ ), every the weighted mean of values of this function is greater than the values of the same function for weighted mean values of the variable:

$$\widetilde{u \log u} \geq \tilde{u} \log \tilde{u}$$

In the present case we have, with  $u$  replaced by  $P(M)$ ,

$$\widetilde{u \log u} = \iint Q(M, M') P(M) \log (M) dM$$

$$\tilde{u} \log \tilde{u} = P'(M') \log P'(M')$$

Substituting these in the above convexity inequality, and then integrating with respect to  $M'$  over  $\bar{\Sigma}$ , with application of (15.3), gives the desired conclusion (15.6). A more detailed study of convexity methods gives the last part of the theorem. See Appendix G for more details.

Theorem 3. The minimum value of  $G[P(M)]$  occurs when  $P(M)$  is the uniform distribution  $P(M) = 1/\text{area } \bar{\Sigma} = 1/\bar{\Sigma}$ .

Information-minimizing expressions under various constraints (e.g., given moments) are usually sought by the formal methods of the calculus of variations. When such minimizing distributions exist (which is the exception rather than the rule), and when an explicit formula has been found for them, the fact that they actually have the minimum property is often easily established by elementary relations of convexity. In the present case we wish to show that for every distribution  $P(M)$ , the uniform ones gives a  $\leq$  informational integral; i.e. that

$$(15.7) \quad \iint P(M) \log P(M) dM - \iint \frac{1}{\bar{\Sigma}} \log \frac{1}{\bar{\Sigma}} dM \geq 0$$

On setting  $P^*(M) = P(M) \bar{\Sigma}$  (which is the ratio of the given to the uniform probability density) it is seen at once that the above

difference can be written as

$$\iint P^*(M) \log P^*(M) \, dM/\bar{\Sigma} = \overline{P^* \log P^*}$$

which is a mean using  $1/\bar{\Sigma}$  as weighting factor. It is therefore not less than  $\overline{P^*} \log \overline{P^*}$ , by the convexity; and since

$$\overline{P^*} = \iint \bar{\Sigma} \cdot P(M) \, dM/\bar{\Sigma} = 1$$

The validity of (15.7) follows--actually with  $(\geq)$  replaced by  $(>)$  when  $P(M)$  is not uniform.

From this result the formulas for the acoustical intensity at long ranges can be obtained in simple form, involving areas of the parts of the surface of section relevant to the positions of the sound source and the receiver. Before giving them, two remarks are in order:

First, the above results are independent of the coordinate system chosen on the surface of section. We could equally well have chosen  $(z, \gamma)$  rather than  $(z, p_z)$  as noted in Section 14; then, not areas, but masses of density  $p_r$  would have been the representatives of the basic phase integral invariant. The corresponding density factor would have appeared in the definition of the informational integral, in which  $dM$  would be  $p_r \, dz \, d\gamma$  instead of  $dz dp_z$ , etc. The steady distribution approached would not be uniform with respect to unit area in the  $zy$ -plane, but only with respect to unit mass--with the above density. Similarly if other coordinates are used in  $\bar{\Sigma}$ .



Second, unless there is an integral invariant of the stochastic transformations on  $\bar{\Sigma}$ , (15.7) in Theorem 2 is false: we cannot assume that the minimization of  $G[P(M)]$  has any bearing whatsoever on our physical problem. Counter-examples to Theorem 2 are, in fact, easily given, since stochastic transformations of the type (15.2) could change a uniform distribution into a non-uniform one--whether with respect to unit area or to unit mass. It is necessary to emphasize this state of affairs because various attempts are often made to get an answer without the necessary complete hypothesis of an integral invariant. More recent developments of information theory have been based on the fact that, in last analysis, this concept is a property of pairs of distributions<sup>( )</sup>, making possible correct physical applications of a minimum principle. In the present case, our two distributions are  $P(M)$  and the Hamiltonian integral invariant.

We turn now to the application of these theorems to the evolution, with increasing  $r$ , of the power flow density  $g(r, z, p_z) = g(r, M)$  across  $\bar{\Sigma}(r)$  (per unit azimuth).

Up to now everything has been based on the notion that this quantity is determinate--its value being that of  $g(r_0, M_0)$ , where  $M_0$  and  $M$  lie on the same ray, determined by the given differential equations. Now, however, the latter equations are not unique, depending as they do on the particular choice of  $V$  from the ensemble  $\{V\}$ . Accordingly, the density  $g(r, M)$  is a chance variable: it could be zero if the ray

through it does not cut  $I(r_0)$ ; or  $q$  if it does. What then do we mean by this power flow density? Once the question is put in this form, the answer seems obvious: the power flow density at  $(r, M)$  is the mean (or expected value)  $\bar{g}(r, M)$  of the random variable  $g(r, M)$ , over the "population"  $\{V\}$ . Once this conception is adopted, the law of evolution of the density with increasing  $r$  is easy to formulate.

We have defined the "stochastic kernel"  $Q(M, M') = Q(M, r; M', r')$  as the probability density of points  $M'$  of intersection with  $\bar{\Sigma}(r)$  at  $M$ :  $Q(M, M')dM'$  is the conditional probability that a ray given to pass through  $M$  at range  $r$ , shall cut  $\Sigma(r')$  at a point in the elementary region  $dM'$  containing  $M'$ . What we need is the inverse probability,  $\check{Q}(M, M')$ , where  $\check{Q}(M, M')dM$  is the conditional probability that a ray cutting  $\bar{\Sigma}(r')$  at  $M'$  did cut  $\bar{\Sigma}(r)$  at a point in  $dM$  (containing  $M$ ). This is given by Bayes' theorem on inverse probability:

$$\check{Q}(M, M') = \pi(M)Q(M, M') / \iint \pi(M)Q(M, M') dM,$$

where  $\pi(M)$  is the a priori probability density of positions of the point  $M$  on  $\bar{\Sigma}(r)$ :

$$\pi(M) dM = \text{prob.} \{ M \text{ in } dM \text{ on } \bar{\Sigma}(r) \}$$

This density  $\pi(M)$  is unknown; however, being a probability distribution on  $\bar{\Sigma}(r)$ , it evolves with increase of  $r$  in accordance with (15.2)—i.e., replacing  $P$  by  $\pi$ . Therefore as the ranges increase it will approach the value of least G-information; i.e., a constant.

Then Bayes' formula, combined with Theorem 1, shows that  $\hat{Q}(M, M') = Q(M, M')$ . Stated in common language, this means that the transition probability  $M' \rightarrow M$  equals the transition probability  $M \rightarrow M'$ .

Suppose first that the actual (deterministic) values of  $g(r, M)$  are given on  $\bar{\Sigma}(r)$  thus determining the set  $\bar{\Sigma}(r, u, \Delta u)$  of points for which  $g(r, M)$  is between  $u$  and  $u + \Delta u$ . The probability that  $g(r', M')$  lie in the same interval is that the ray through  $(r', M')$  cut  $\bar{\Sigma}(r)$  in the above set; viz.,

$$\iint_{\bar{\Sigma}(r, u, \Delta u)} Q(M, M') dM .$$

If the interval of all possible values of  $g$  (namely, from 0 to  $q$ ) is subdivided into a large number of small pieces such as  $\Delta u$ , and the sum of corresponding integrals of the above form are multiplied by the corresponding  $u$  values and added, a sum is obtained which, as the  $\Delta u \rightarrow 0$ , approaches the required expected value  $\bar{g}(r', M')$ . On the other hand, the sum clearly approaches the integral of  $g(r, M)Q(M, M')dM$  over  $\bar{\Sigma}(r)$ . Thus

$$\bar{g}(r', M') = \iint g(r, M) Q(M, M') dM .$$

This has been established on the supposition that  $g(r, M)$  is a given function on  $\bar{\Sigma}(r)$ . If, on the contrary, it is the random function corresponding to the random choice of  $V$  from  $\{V\}$ , only  $\bar{g}(r, M)$  can be known. But the above formula, being linear in the random quantity  $g$ , can be applied to the mean, thus leading to the formula:

$$(15.8) \quad \bar{g}(r', M') = \iint \bar{g}(r, M) Q(M, M') dM$$

But this is identical with (15.2) with  $P'(M')$  and  $P(M)$  replaced by  $\bar{g}(r', M')$  and  $\bar{g}(r, M)$ . Suppose that we divide (15.8) through by the common value ( $P_2$ ) of the integrals of these two functions over  $\bar{\Sigma}$ : the resulting densities will have exactly the same mathematical properties as  $P'(M')$  and  $P(M)$  do, namely, non-negativeness, integrating to unity over  $\bar{\Sigma}$ , and evolving (at long range) by the formula (15.2). Consequently the logarithmic integral expression  $G[\hat{g}(r, M)/P_2]$ --while no longer the "information"--obeys the purely mathematical theorems established for  $P(M)$ : it approaches a constant.

Thus the intuitive picture of diffusive mixing in a liquid is justified. But our methods actually allow a more general result to be established--getting rid of the restriction that on the initial injection surface  $I$  of  $\Sigma(r_0)$ ,  $g(r_0, z, p_z) = q$ ,  $= 0$  off  $I$ . For the equation (15.8) does not depend on this assumption: all that is needed is that this initial function  $g$  be non-negative and integrate to the total power injected across  $I$ , per unit azimuth. The same uniform spread at large ranges across  $\bar{\Sigma}$  follows by exactly the same reasoning.

On the other hand, Assumption 2 of Section 13 ( $\bar{V} = U(z)$ : quazilaminarity) is essential to Sections 13, 14, 15. The case of gradual and non-random departures from laminarity, while treated numerically by

perturbation methods in the companion study (referenced at the end of Section 14), and coming under the general considerations of Sections 3 through 12, is not examined in terms of special theorems up to this point.

We note in closing that the factor  $e^{-\alpha r}$  expressing the random "leakage" out of  $\bar{\Sigma}$  has been methodically left out from the above formulas and must be restored in the final expression of transmission loss.

## 16. Transmission Loss with Secular Non-Laminarity

The situation of importance to intermediate and long range acoustic reception is not one of strict laminarity  $c = c(z)$ --either when  $c$  is the given sound speed, as in Section 14, or when determined by averaging, as in Section 15. The actual situation, as shown by masses of oceanographic observations in many oceans and seas and at various seasons, is the one in which the sound speed  $c$  (and hence  $n = 1/c$  and  $V = -n^2/2$ ) vary not only with depth  $z$ , but also horizontally--with geographical position on the ocean. If this horizontal variation were anywhere nearly as great as the variation with depth, no effective sound channel would exist, since the acoustic power of any signal would be too much refracted out of the channel, and would therefore be undetectable at the ranges of present concern. What does occur, as the oceanographic measurements indicate, is a slight but progressive change in acoustic profile over scores or hundreds of miles, a variation that has a cumulative effect on the transmission loss. This situation we shall term secular non-laminarity. More precise characterizations of this condition and of its effects on propagation loss calculations will be provided during the course of the mathematical developments now to be presented, in this special case of "quasi-laminarity".

The directly practical object of the present section is to extend the transmission loss equations of Section 14, e.g., (14.19) and (14.20), to the case of secular non-laminarity, expressing the results in terms of the supposedly known acoustic profiles at two points: at the emitter and at the receiver. This will provide a theoretical basis

for the formulas and schemata of numerical computation contained in the correlative study of the present series, referenced at the close of Section 14: "Computation of Long Range Propagation Losses in a Duct" (June 20, 1973). The more indirectly practical object of the present section is the formulation of an approach to the calculation of transmission loss under far more general conditions of secular non-laminarity, with  $c = c(x, y, z)$ , the rays being twisted space-curves.

It is relevant to our subject to note that no mathematical treatment\* seems to be available by which transmission loss can be computed under the conditions of secular non-laminarity assumed in this section, and going as far as the one in the companion study referenced above--other than those based on the ray-by-ray constructions requiring massive quantities of specific data and much time on computing machines of very great capacity, and using unrealistically precise knowledge of oceanographic fine-structure.

The present method rests on two bases: a geometrical property of integral surfaces of the ray equations in phase space; and a statistical assumption of ray incoherence (implied in Section 14 and outlined in the correlative study referred to above) allowing approximation by averages.

At the risk of repetitiveness, three essentially equivalent treatments, of increasing levels of mathematical technicality, will be given: first, an intuitive extension of the laminar methods of Section 14; second, a treatment based on the geometric interpretation of Poisson bracket relations; finally, a general mathematical approach requiring new adaptations of classical theorems, and capable of extension to very general cases.

---

\* We are excepting derivations of formulas equivalent to the 2-profile one referenced above by some authors who confine themselves mainly to physical intuition and formal analogy, without a rigorous derivation from precisely stated physical assumptions.

While of more general applicability, we shall for simplicity of presentation confine the study to the case of azimuthal symmetry, with the rays in vertical planes through an emitter reference point, and apply the cylindrical coordinates and other general assumptions of Sections 13 and 14. Now our "potential"  $V = -n^2/2 = -1/2c^2$  depends on both  $r$  and  $z$ . The ray differential equations (13.5) in the phase space  $\mathfrak{S}_3$  may be written in the symmetrical form as follows (subscripts of  $V$  denoting partial derivatives):

$$(16.1) \quad \frac{dr}{p_r} = \frac{dz}{p_z} = -\frac{dp_r}{V_r} = -\frac{dp_z}{V_z} = d\tau$$

As explained before,  $\mathfrak{S}_3$  is the locus of the equation

$$(16.2) \quad H = \frac{1}{2} (p_r^2 + p_z^2) + V = 0$$

in which locus each ray lies, one and only one through each point  $(r, z, p_z)$ ; see Figures 14.I and II. Equations (13.6) which we now write in terms of the angle of inclination  $\alpha$  of the tangent to each ray with the horizontal, show that

$$(16.3) \quad p_r = n \cos \alpha, \quad p_z = n \sin \alpha, \quad dz/dr = \tan \alpha$$

The geometry of integral surfaces. We shall apply to the points  $(r, z, p_z)$  of  $\mathfrak{S}_3$  some of the general elementary considerations of the early part of Section 9: in particular, the notion of the "first integral"  $F$  which satisfies the partial differential equation



(9.4) related to the system of ordinary differential equations (9.1). In the present 3-dimensional case, the latter are of course (16.1) in which the fourth variable (e.g.,  $p_r$ ) can be expressed in terms of the other three ( $r, z, p_z$ ) by means of (16.2). If this implicit method seems to complicate the picture, we can of course use (16.3) to eliminate both  $p_r$  and  $p_z$ , thus having the three variables ( $r, z, \alpha$ ) as coordinates of points in  $\mathcal{S}_3$ --unconnected by any equation, merely satisfying the simple inequalities of obvious physical meaning (c.f. (14.1), (14.2)):

$$(16.4) \quad r \geq 0; \quad a \leq z \leq b; \quad -\pi/2 < \alpha < \pi/2.$$

The partial differential equation for a first integral  $F = F(r, z, p_r, p_z)$  is (using subscripts for its partial derivatives):

$$(16.5) \quad p_r F_r + p_z F_z - V_r F_{p_r} - V_z F_{p_z} = 0.$$

In the variables ( $r, z, \alpha$ ) the equations (16.1) and (16.5) become, if we write  $m = \log n = \frac{1}{2} \log (-V) + \frac{1}{2} \log 2$ :

$$(16.6) \quad \frac{dr}{\cos \alpha} = \frac{dz}{\sin \alpha} = \frac{-d\alpha}{m_r \sin \alpha - m_z \cos \alpha} = nd\tau$$

$$(16.7) \quad \cos \alpha F_r + \sin \alpha F_z - (m_r \sin \alpha - m_z \cos \alpha) F_\alpha = 0$$

while (16.4) still applies, but (16.3) is redundant. It turns out that the formulation (16.1) through (16.5) is more appropriate to our purposes than (16.6) and (16.7), which will not be used in our later investigations.

Equations (16.5) and (16.7) are of course the expression of the fact that the differential  $dF$  along the rays (solutions, or integral curves) determined by (16.1) or (16.6) is zero; equivalently, that  $F$  remains constant along all such curves. It follows that if  $C$  is any constant for which the locus  $F = C$  is in  $\mathfrak{P}_3$ , every ray either lies wholly on this locus, or has no point in common with it. Thus we can state that the locus  $F = C$  is made up of rays: in general it is a surface swept out by all the rays through some curve in  $\mathfrak{S}_3$ . Such a surface is an integral surface. Not only is every integral surface made up of rays, but the intersection of a pair of integral surfaces is a ray (or a set of rays, usually distinct): for evidently the ray through any point common to the two integral surfaces lies in each, and hence, in their intersection.

Finally, there can be no more than two functionally independent (i.e., essentially different) first integrals. Thus if  $(F_1, F_2, F_3)$  is a set of such integrals and if every  $F_1 = C_1, F_2 = C_2$ , is a single integral curve (the "general" case),  $F_3$  must be constant along it; therefore when the values of  $F_1$  and  $F_2$  are given, that of  $F_3$  is determined: by definition of "function", it is a function of the two former,

$$(16.8) \quad F_3 = \Omega (F_1, F_2).$$

Conversely, if  $F_1$  and  $F_2$  are first integrals and  $\Omega$  is any function of two variables, the  $F_3$  as defined by (16.8) is a first integral.

We add to the above set of elementary facts on differential equations the further ones derived from the existence and uniqueness theorems in the same field--stated after  $r$  has been introduced in (16.1) or (16.6) as the independent variable instead of  $\tau$  (which plays no part in the geometry). Let  $(z^0, p_z^0)$  be the values assumed by  $z$  and  $p_z$  when  $r = r^0$ : "initial values", in which  $r^0$  is to be a small range close to the emitter;  $(z^0, p_z^0)$  are the coordinates of the surface of section  $\Sigma(r^0)$  defined in Section 14. Through the "initial point"  $(r^0, z^0, p_z^0)$  passes just one integral curve (ray in  $\mathfrak{S}_3$ ), along which we have

$$(16.9) \quad \begin{aligned} z &= Z(r; r^0, z^0, p_z^0) \\ p_z &= P_z(r; r^0, z^0, p_z^0). \end{aligned}$$

Since the letters with superscript zero are the initial values, when  $r$  is set equal to  $r^0$  these two functions reduce to  $z^0$  and  $p_z^0$  respectively. Moreover, the roles of the plain and the zero-superscript variables may be interchanged: the equations remain valid:

$$(16.10) \quad \begin{aligned} z^0 &= Z(r^0; r, z, p_z) \\ p_z^0 &= P_z(r^0; r, z, p_z) \end{aligned}$$

This is the statement in formulas of the property of the (unique) integral curve through one point and containing a second being the one determined (uniquely) by the second. Of course one tacit assumption has been made: that a duct actually exists and that we are confining

---

\* Upper zeros are more convenient here than the lower zeros used before.

our attention to the part of  $\mathcal{S}_3$  composed of the rays remaining in the duct throughout the ranges considered (cf. the "central core" discussed in Section 14).

Equations (16.10) automatically give us two independent first integrals of our differential system: we may suppose a numerical value given once and for all to  $r^0$  (e.g., zero; or a nominal range taken as the unit range from the emitter). Then we may write the right-hand members as  $Z(r, z, p_z)$  and  $P_z(r, z, p_z)$ . That each remains constant along every integral curve is the evident consequence of the fact that the coordinates of a ray's intersection with the surface  $\Sigma(r^0)$  are evidently determined by the ray itself and do not change as we go along it. This reasoning is general and shows how the  $n$ 'th order differential equations discussed in Section 9 always have  $n-1$  independent first integrals.

In view of the geometrical properties of our integral surfaces, we might hope to throw light on the integral curves by examining the surfaces  $Z = \text{constant}$  and  $P_z = \text{constant}$ . This turns out, however, to be an impracticable idea because of the extraordinary complexity of these loci, once one or two convergence zones have been reached: this is not the result of the non-laminarity assumed here, but the nature of families of rays in a duct.

To understand this fact, let us assume the laminar case and attempt to visualize, in the  $\mathcal{S}_3$  shown in Figure 14.I, the locus of  $Z(r, z, p_z) = \text{constant} = z^0$ . It is generated by the rays which cut  $\Sigma(r^0)$  in a horizontal line (depth  $z^0$ ): along this line the rays have a

continuous infinitude of different values of the Snell constant  $k = p_r^0$ , and hence lie on different mutually nested cylindrical surfaces (only one is shown in Figure 14.I). Since the rays on each cylinder wind about it in helix fashion, but with a mean pitch (length of ray period) which is different on different cylinders, and in fact varies considerably with  $k$ , the complexity of the surface swept out by these variable pitch twisted rays can be well imagined. It was indeed this rapidly developing complexity that justified our assumptions of "ergodic mixing" of regions on the surface of Section  $\Sigma(r)$  with increasing  $r$ , and hence the simplifying limiting methods applied in Section 14--and to be used again in a more general form here. Exactly as complex is the structure of the locus of  $P_z(r, z, p_z) = \text{constant} = p_z^0$ : it is swept out by the rays through the corresponding vertical line segment in  $\Sigma(r_0)$ .

It is at this point that the special significance of the further first integral,

$$(16.11) \quad p_r^0 = P_r(r^0; r, z, p_z),$$

which we will write  $P_r(r, z, p_z)$ , becomes apparent. [Of course it is not independent of the other two, and is expressed in terms of them by use of (16.2); c.f. (16.12)]. The point is that in the laminar case  $P_r(r, z, p_z) \equiv p_r$  (the constancy of which expresses Snell's law): the locus in this case of each equation  $P_r = \text{constant} = p_r^0$  is a cylindrical surface in  $\mathfrak{H}_3$ --a figure which even at very long ranges is "simple".

If, then, the choice of integral  $P_r$  gives simple surfaces--in contrast to  $Z$  or  $P_z$ --in the laminar case, how does the situation stand in the case of secular non-laminarity? (We re-emphasize that the property of being first integrals is possessed by the  $Z$ ,  $P_z$ ,  $P_r$ , defined by (16.10) and (16.11), in the non-laminar case as well as in the laminar).

The answer to the above question is contained in a set of general theorems in differential equations that complement the classical existence and uniqueness theorems, and are due to the work early in this century of É. Picard, H. Poincaré and G. D. Birkhoff; they are brought together in all clarity of elementary detail in the first chapter of G. D. Birkhoff's "Dynamical Systems" (American Mathematical Society Colloquium Publications, Volume IX, 1927). That reference is concerned with the way in which the solutions such as (16.9) depend on the initial conditions. Also, in the case that the coefficients of the differential equations, such as (16.1) or (16.6), vary--e.g., when they contain continuously varying parameters--the effect of such variation on the solutions is established. It is shown, in particular, that slight changes in coefficients produce slight changes in any given first integral ("given", e.g., by its initial values, on  $\Sigma(r^0)$ ). Moreover, the "smoothness" (order of continuous derivations) in the solutions will match that of the continuously varying coefficients--going to the point of analyticity when the latter are analytic. This is established, of course, for a finite, but arbitrarily large, range: corresponding with the physical reality of the situation studied here.

To apply these results to the case of secular non-laminarity, let us take a fixed range, such as  $r^0$ , and set  $U(z) = V(r^0, z)$ ,  $W(r, z) = V(r, z) - U(z)$ . Then consider the family of differential equations obtained by replacing  $V$  in (16.1) or (16.6) by  $U(z) + \epsilon W(r, z)$ , where  $\epsilon$  is a continuous parameter: as it goes from 0 to 1, the differential equations go from the laminar approximation to the non-laminar actuality. Since we are assuming secularity in the latter, the change in coefficients is slight, even when  $\epsilon = 1$ --in a sense to be made precise below.

From the theorems cited, during this continuous increase in  $\epsilon$  the solutions (16.9), as well as the integrals  $Z$ ,  $P_z$  and  $P_r$  (which now are functions of  $\epsilon, r, z, p_z$ ) change continuously: a point at which an integral curve cuts any given  $\Sigma(r)$  changes only slightly if  $\epsilon$  changes slightly; and similarly for the intersections with  $\Sigma(r)$  of any of the integral surfaces. But there is this fundamental difference in the three latter cases: whereas the integral surfaces  $Z = z^0$  and  $P_z = p_z^0$  ( $\epsilon=0$ ) were excessively complicated and had intersections running all through  $\Sigma(r)$ , so that a slight change will not change this feature--the intersections with  $P_r = k$ , on the other hand, are the slight deformations of simple closed curves (those shown in Figures 14.III and V) and hence will remain simple and closed. This will cease only if  $r$  is so large that, even with the moderate change in coefficients produced by the above change in  $\epsilon$  away from zero, the cumulative effect of increase of range is too great. In this connection we can state the following mathematically more precise rendering of the first condition for "secular non-laminarity":

The first part of our definition of secular non-laminarity in the range interval  $(r^0, r^1)$  that as  $\epsilon$  goes from 0 to 1, the intersections with  $\Sigma(r)$  of the integral surfaces  $P_r = k$  remain simple closed curves, cut in no more than two points by any line  $z = z^0$ .

This will mean that these integral surfaces remain tubular and nested; that they each have a locus on them of highest (shallowest) and lowest (deepest) points,  $z = z_1(r)$  and  $z = z_2(r)$  respectively; and that for every intermediate depth  $[z_1(r) < z < z_2(r)]$  just two values of  $p_z$  correspond. Of course the integral curves will wind about these tubular "deforms" of the laminar cylinders as they did when  $\epsilon = 0$ . In other words, the essential geometrical basis for the work on "ergodic mixing" and limiting behavior of the power flux, expounded in Section 14, can be carried over to its secular approximation. We re-emphasize that the preservation of areas, being the result of the general Hamiltonian property, is valid in all cases--and so, accordingly, is the fact that the power flux density quantity  $g = p_r \tilde{r} f$  of (14.4) is itself a first integral, as shown in the reasoning of Section 13 leading to (13.11). There is of course always the tacit assumption that our constants are chosen so that the rays stay in the duct.

At this stage we have four first integrals:  $Z, P_z, P_r, g$ , all functions of  $(r, z, p_z)$  or, equivalently, of  $(r, z, \alpha)$ : by general principles, only two can be independent. The first three are connected by the relation

$$(16.12) \quad P_r^2(r, z, p_z) + P_z^2(r, z, p_z) = n^2(r^0, Z(r, z, p_z))$$



which is an obvious consequence of (16.10) and (16.11), applied to (16.2) (writing  $V = -n^2/2$ ). On the other hand, beyond the general fact that  $g$  must be some function of an independent pair chosen from the first three of the above (as in (16.8)), we know nothing. We can, however, be sure of one thing: it is either very "complicated"--i.e., the loci  $g = \text{constant}$  are complicated integral surfaces--or else it is a function of  $P_r$  only. This is because otherwise the intersections of  $g = \text{constant}$  with  $P_r = \text{constant}$  would be "simple" and could not wind about the "tubes" given by the latter. This heuristic argument is a sort of repetition of the ideas expounded in Section 14 with the aid of Figure 14.V, and leading to the ergodic mixing and equations (14.6) and  $h(z, p_z) = F(k)$  in the laminar case.

The intuitive approach. We are now in a position to parallel, in the present case, the reasoning applied to the laminar case in Section 14, which led from (13.10) to the final expressions (14.19) and (14.20) for the transmission loss in that case. To begin with, we emphasize that the reasoning of Section 13 which established (13.10) was based solely on the general assumption involving azimuthal symmetry, and not on that of laminarity: this allows the use of (13.10) as our present starting point. Furthermore, we shall assume that over the initial injection region  $I$  near the source ( $I = I(r^0)$ ) the power flux density (per unit  $zp_z$  - area) is constant: equivalently,  $g$  is the positive constant  $q$ ; while  $g = 0$  at points of  $\Sigma(r^0)$  not on  $I$ . This assumption has the same reasons for validity as in the laminar case. Introducing the resulting expression for  $\tilde{f}(r^0, z^0, p_z^0)$  into (13.10), we obtain,

exactly as in the earlier case, the equation (14.8) for the acoustic intensity over  $I$  at the source in terms of the constant  $q$ .

On the other hand, the reasoning leading to (14.6) and  $h(z, p_z) = F(k)$ , and thence to the replacement of  $\tilde{f}$  by  $\tilde{f}$ , with the eventual derivation of (14.9) and (14.10), has to be re-stated in modified form, for the three following reasons: first, as noted before, the integral surfaces  $P_r = k$  are no longer horizontal cylinders, but irregular tubes running in a general way along the duct; second, we cannot without too great a stretch of realism speak of "limits as  $r \rightarrow \infty$ "--only of "acceptable approximations for the large values of  $r$  encountered in the type of acoustic detection under study"; third, the ergodic theorems can no longer serve us, as they did in the laminar case, since in their proof, the group property (of the transformations on  $\Sigma(r)$ ) is conventionally assumed: as noted in Section 15, this is absent in the non-laminar case.\* We shall first indicate on an intuitive basis the ideas which both parallel and modify those applied to the laminar case; and after this orientation, pass on to their more abstract mathematical validation.

If we think of the thin horizontal strip  $I = I(r^0)$  shown in Figure 14.V as dyed black, and all the integral curves that cut it as carrying this black stain, we see that every region  $I(r)$  in which they intersect  $\Sigma(r)$  will likewise have this stain--all as in the earlier case. Furthermore, by the area invariance of the transformations on the surface of section  $\Sigma(r)$  with increasing  $r$ , the stained region  $I(r)$  will

---

\* To reappear after the canonical deformation applied below.

the same  $zp_z$ -area as  $I$ --and so will the image in  $I(r)$  of any piece of  $I$ . As in the earlier case, the curves  $P_r = k$  on  $\Sigma(r)$  are tangent to the paths of the points on  $\Sigma(r)$  as  $r$  increases; but in the present case they vary with  $r$ . Nevertheless, as they move, they form a moving boundary of the stained regions: in particular, the total stained area in the (irregular and varying) ring on  $\Sigma(r)$  for which  $P_r$  has values between  $k$  and  $k + \Delta k$  does not change with  $r$ . We are particularly interested in this ring when  $\Delta k/k \ll 1$ . In the present case the rays are not strictly periodic; but if we define their approximate period as the horizontal range interval from one point of greatest depth to the next one, we can say that this "pseudo-period" is very dependent on the "pseudo-Snell constant"  $k$  of the integral surface  $P_r = k$  upon which they lie. Consequently the advance along their paths in  $\Sigma(r)$  will be different for the different paths, so that the picture of "contorsion" and "mixing" suggested in Figure 14.V will apply. This has the following consequence, corresponding in the present case to what was expressed by (14.6) etc. in the earlier one:

Let  $\Delta\Sigma$  be a "simple" piece of the ring of  $\Sigma(r)$  in which  $P_r$  is between  $k$  and  $k + \Delta k$ , and let  $\Delta I(r) = I(r) \cap \Delta\Sigma$  be the area of this piece that is stained by the dye injected into  $I$ . Then the ratio of areas  $\Delta I(r)/\Delta\Sigma$  is approximately independent of the position and (simple) shape of  $\Delta\Sigma$  on the ring--once the range  $r$  exceeds a few convergence zones. This is the present form of the earlier relation (14.6); but it requires an explanation: the stipulation that  $\Delta\Sigma$  be a "simple" piece of the ring is intended to exclude constructions that could only

be formulated after the exact shape of  $I(r)$  became known. Examples of simple pieces would be the parts in the ring of a rectangular subdivision of  $\Sigma(r)$ ; or a piece bounded by two curves of equations  $P_r = \text{constant}$  and two segments of the normals to one cut off by the other; etc. The sufficiency of confining certain later constructions to the use of such simple pieces appears below. As in Section 14 it is supposed that  $\Delta\Sigma$ , although always on  $\Sigma(r)$ , does not "flow" as  $r$  increases.

From the above assumption, we may calculate the common ratio: let the ring ( $P_r$  between  $k$  and  $k + \Delta k$ ) be subdivided into  $n$  simple pieces of equal area,  $\Delta\Sigma_i$ ; the corresponding areas they cut from  $I(r)$  will be equal, since according to the above assumption  $\Delta I(r)_i / \Delta\Sigma_i$  is the same for all  $i = 1, \dots, n$ . It follows at once that this common value is equal to

$$\frac{\text{area of the part of } I(r) \text{ in ring } (k, k + \Delta k)}{\text{total area of the ring } (k, k + \Delta k)}$$

This is, therefore, the value of the ratio of any  $\Delta I(r)\Delta\Sigma$ , whatever the size of the (simple) sub-region  $\Delta\Sigma$  of the ring may be. This conclusion has the following consequences:

First, by the invariance of areas in the transformation on the surface of section  $\Sigma(r)$ , and the fact that  $P_r = k$  is an integral surface, the above ratio is independent of  $r$ : it depends on  $k$  and  $\Delta k$  only.

Second, from the fact that to quantities of the first order in  $\Delta k$ , the numerator and denominator in the above fraction are proportional to  $\Delta k$  (the denominator, by the properties of areas; the numerator also,-- once our assumption has been applied to its pieces) it follows that our fraction is (approximately, for small  $\Delta k$ ) a function of  $k$  only. Therefore

$$(16.13) \quad \frac{\Delta I(r)}{\Delta \Sigma} \approx \frac{\text{area of } I(r) \text{ in ring } (k, k + \Delta k)}{\text{area of ring } (k, k + \Delta k)} \approx F(k)$$

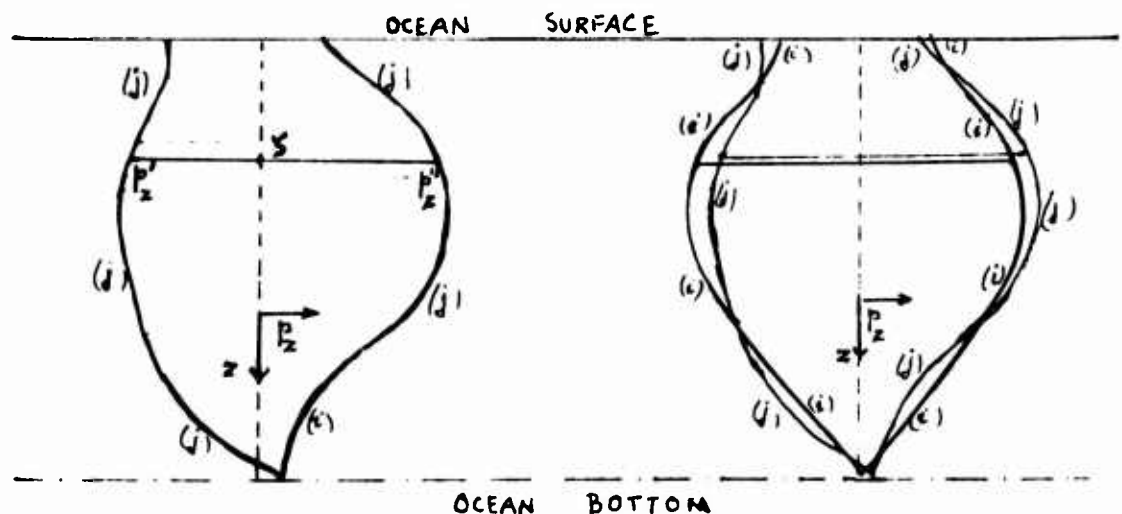
Third, because of the independence of the quantities in (16.13) of range  $r$ , they can be calculated at  $r = r^0$ . But there  $\Sigma = \Sigma(r^0)$  and  $I = I(r^0)$  have the form shown in Figure 14.V, from the geometry of which we derive equation (14.15) for  $F(k)$  and therefore (in the case described subsequently to that formula) the expression (14.17). Of course  $R(k)$  is the "pseudo-period" at the emitter; but it is still given by (14.16).

Fourth, instead of the power flux density  $\tilde{f} = \tilde{f}(r, z, p_z)$  across  $\Sigma(r)$ , we can use, in (13.10), the function  $\tilde{\tilde{f}}(r, z, p_z) = (q/rp_r)F(k)$ , just as we did in deriving (14.9); this will lead to (14.10) and similar consequences; but there are certain differences:

Let us first recall briefly the reasoning used before in the replacement of  $\tilde{f}$  by  $\tilde{\tilde{f}}$ : by definition,  $\tilde{\tilde{f}}$  when integrated over any piece  $\Delta \Sigma$  of  $\Sigma(r)$  gives as an answer the total power flow through that piece; and this is strictly equal to the area  $\Delta I(r)$ , times the value of  $\tilde{f}^0$  given in terms of  $q$  by the equation  $r^0 p_r^0 \tilde{f}^0 = q$  assumed in the first

paragraph of our intuitive approach, as well as in Section 14 in the derivation of (14.8). On the other hand, the ratio of  $\Delta I(r)/\Delta \Sigma$  being approximately  $F(k)$  at long ranges, in virtue of (16.13), it is seen that we get the same results at such ranges by the replacement of  $\tilde{f}$  by  $\tilde{f}^*$  as described above. We are therefore justified in the use in the present case of (14.10) as well as its immediate consequence (14.10)' as expressing the acoustic intensity (power flux per unit surface crossed by the acoustic power). The new feature is in the evaluation of these integrals at the range  $r^1$  of the receiver.

Figure 16.I Surface of Section at Receiver



(a) Limiting Pseudo-Snell  
Curve (j)(j)(j) . . . .  
 $P_z = k$

(b) Limiting Curve (j)(j)(j) . . .  
with superposed mean Snell  
Curve (i)(i)(i) . . .  
 $P_z = k'; r^1 < r < r^2$

The geometry of the situation is illustrated schematically in Figure 16.Ia which shows the curves  $P_r(r^1, z, p_z) = k$  on  $\Sigma_1 = \Sigma(r^1)$ ; it may be regarded as the result of a moderate deformation of Figure 14.III. While still cutting a single segment from each horizontal line ( $z = \text{constant}$ ), it is no longer precisely symmetric about the vertical  $z$  - axis.

Suppose now that we denote the effective depth interval of the receiver by  $(\zeta', \zeta'')$ . To find its angular interval  $\Gamma_1$ , or equivalently, the interval of  $p_z$  at the receiver, we note that this is the horizontal segment cut off from the horizontal line  $z = \zeta$  (a depth between  $\zeta'$  and  $\zeta''$ ) from the curve  $P_r(r^1, z, p_z) = k$ , where  $k$  has its extreme value of  $k$  for all the rays in the duct. Actually, it is the minimum  $k$ , since these tubes expand as  $k$  decreases. Strictly speaking, this value will depend on the relation between the variations of both the acoustic profile and the bottom throughout the range: it is the least value of  $k$  for which the tubular integral surface  $P_r = k$  is tangent to the bottom. On the assumption that the bottom is essentially flat and that there is little appreciable variation in the sound speed along it, we may accept the limiting  $k$  found at the source, where  $k = P_r = p_r^0 = n(r^0, b)$ . Therefore the  $p_z$ -interval is bounded by the two values of  $p_z$  satisfying

$$(16.14) \quad P_r(r^1, \zeta, p_z) = n(r^0, b).$$

They are denoted by  $p_z'$  and  $p_z''$  in Figure 16.Ia: let the corresponding angles be  $\gamma'$  and  $\gamma''$ . Thus  $\Gamma_1: (\gamma', \gamma'')$  and the two components of our power flux are given by (14.10') with

$$(16.15) \quad Z = \frac{q}{r} \int_{\gamma'}^{\gamma''} F(k) n \cos \gamma d \gamma$$

$$R = \frac{q}{r} \int_{\gamma'}^{\gamma''} F(k) n \sin \gamma d \gamma$$

where  $k$  is the function of the variable of integration  $\gamma$  determined by the implicit equations.

$$(16.16) \quad P_r(r^1, \zeta, p_z) = k, \quad p_z = n(r^1, \zeta) \cos \gamma$$

Since, as we have shown,  $F(k)$  is still given by (14.7), the only remaining task in evaluating the transmission loss is to find a satisfactory approximation to the limits of integration in (16.15), determined by (16.16).

At the present intuitive level, we may reason as follows: think of a fictitious medium of propagation, coinciding with the actual one in the range interval  $(r^0, r^1)$ , but whose index  $n(r, z)$  becomes independent of  $r$  in the interval  $(r^1, r^2)$ , throughout which  $n(r, z) = n(r^1, z)$ ; here  $r^2 - r^1$  exceeds all the periods of the laminar case that sets in as soon as  $r > r^1$ . Finally let the junction between the index, before and after the range  $r^1$ , take place smoothly (with continuous derivatives of all orders throughout a thin junction layer). In  $(r^1, r^2)$  we have laminarity so that  $p_r$  is a first integral of the ray equations. Since  $P_r(r, z, p_z)$  continues to be such an integral, the intersections of their corresponding integral surfaces,  $p_r = k'$  and  $P_r = k$ , will be integral curves. Since the



two surfaces do not in general coincide, these curves will wind about each surface in the manner become so familiar: the same will be true of every "proturbance" of the latter over the former; this can be pictured in terms of their intersections with  $\Sigma(r)$  as  $r$  goes from  $r^1$  to  $r^2$ ; see Figure 16.1(b).

Now we shall "associate" pairs of integral surfaces of the two kinds whenever the areas of their intersections with  $\Sigma(r)$  are equal: the conservation of areas in  $\Sigma(r)$  with increase of  $r$  shows that this association is consistent for all  $r$ . Finally, since the extreme curves are determined by the depths  $\underline{a}$  and  $\underline{b}$  of the ocean surface and bottom, and since  $(b-a) \times$  average horizontal section  $p''_z - p'_z$  (the difference of the extreme values of  $p_z$ ) over the depth interval is the area in question, we see that by associating  $p_r = k'$  with  $P_r = k$  by the areas, we have in fact (at least for the extreme curves) associated the pseudo-Snell curve  $P_r = k$  on  $\Sigma(r)$  with the actual Snell curve which, in the above sense, is its "mean": the statistical assumption of ray incoherence mentioned in our fourth paragraph, shall be construed as allowing us, in the calculation of the limits in (16.15), to replace the extreme  $P_r = k$  by its mean  $p_r = k'$ . When there is but negligible variation in depth  $\underline{b}$  throughout the whole range  $(r^0, r^2)$ , we may readily see that for both extreme curves,  $k'=k$ .

Applying this approximation to (16.15), etc., we are able to complete the derivation of (14.20) for the transmission loss; but emphasizing that the  $z = \zeta$ , the depth of the receiver; that the  $n$  in the limits and integrand is  $n(r^1, z) = n(r^1, \zeta)$  at the receiver; whereas  $R(k)$  is given by (14.16) at the emitter. This completes the validation of (14.20)

and hence of (25.1) in the correlative study. But we wish to go beyond this rather intuitively geometrical treatment. We therefore pass to our second more rigorous treatment, based on the geometric and kinematic properties of the Poisson brackets.

The most convenient starting point is the canonical form (16.1) of the integral curves, restricted by the inequalities (16.4), but considered at first without the condition  $H = 0$ : the rays are the subclass of the integral curves of (16.1) for which  $H = 0$  (so that (16.3) becomes valid). We are dealing, then, with the Hamiltonian system of coordinates  $(r, z)$  and momenta  $(p_r, p_z)$ , defining a 4-dimensional phase space (a 5-dimensional "phase-space-time", if  $\tau$  is included.) The fundamental relative integral invariant of first degree is\*

$$(16.17) \quad \omega(\delta) = p_r \delta r + p_z \delta z - H \delta \tau$$

Now our "simple" integral  $P_r$ , defined as in (16.11), contains  $p_r$ :

$$P_r = P_r(r, z, p_r, p_z).$$

The properties of first integrals, integral surfaces, and integral curves in the case of Hamiltonian systems are most conveniently stated in terms of the Poisson Bracket notation: if  $f$  and  $g$  are any functions of the  $2n$  canonical variables  $q_i$  and  $p_i$  (in our case, four), we write

$$(16.18) \quad [f, g] = \sum_{i=1}^n \left( \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right) \\ = \frac{\partial f}{\partial r} \frac{\partial g}{\partial p_r} + \frac{\partial f}{\partial z} \frac{\partial g}{\partial p_z} - \frac{\partial f}{\partial p_r} \frac{\partial g}{\partial r} - \frac{\partial f}{\partial p_z} \frac{\partial g}{\partial z}$$

---

\* Cf. Appendix E, formulas (E.12), (E.14), (E.16) and (E.17)

Then it is apparent that the canonical equations (E.12) can be written as  $dq_1/dt = [q_1, H]$  and  $dp_1/dt = [p_1, H]$ ; and similarly, the rate of change along any integral curve of a function  $F = F(q, p)$  of the canonical variables\* is  $dF/dt = [F, H]$ . Therefore the condition that  $F$  be a first integral is that  $[F, H] = 0$ . It is obvious from its definition that the Poisson bracket  $[f, g]$  is bilinear: linear in  $f$  and in  $g$  separately (constant coefficients); and that  $[f, f] = 0$  and  $[f, g] = -[g, f]$ . Furthermore, differentiations based on the definition establish the Jacobi identity

$$[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0$$

These simple and well-known† facts lead to some important conclusions. First, if the Hamiltonian  $H$  does not contain the time,  $dH/dt = [H, H] = 0$ , so that  $H$  is a first integral of the differential system which it defines--a fact that we have known and used to select the sub-phase space  $H = 0$  in the case of acoustical rays. Second, if  $F$  is any further first integral of the above system (a function of  $(q, p)$  independent of  $H$  and for which  $[F, H] = 0$ ), it has both a geometric and a kinematic property: the geometric property is the usual one possessed by any first integral and which we have discussed earlier (the loci  $F = \text{constant}$  being fixed in the "flow" determined by the

---

\*This  $F$  has no relation to the earlier  $F = F(k)$ .

†Cf. the exhaustive but not always rigorous treatment given in Chapters XI and XII of E.T. Whittaker "Analytical Dynamics" (Cambridge University Press, 1917). The older notation  $(f, g)$  is used for  $[f, g]$ , the latter being used for the Lagrange Brackets.

differential equation and being made up of its integral curves, etc.)  
 The kinematic property is peculiar to the Hamiltonian nature of the differential equations: let us suppose that we write down a new Hamiltonian system of differential equations, in which  $F$  and not  $H$  is taken as the Hamiltonian; they will have the form--using  $\sigma$  as the "time-like" independent variable--

$$(16.19) \quad dq_1/d\sigma = [q_1, F], \quad dp_1/d\sigma = [p_1, F].$$

They define a continuous 1-parameter family of transformations ("flow") of the  $2n$ -dimensional phase space of  $(q,p)$  into itself. [Since  $F$  does not contain  $\sigma$ , they form a group]. The condition that any function  $G(q,p)$  be a first integral is again that  $[G, F] = 0$ . But since  $F$  was a first integral of the original canonical system based on  $H$ ,  $[F, H] = 0 = -[H, F]$ : therefore  $H$  is also a first integral of the canonical system based on  $F$ ; the loci  $H = \text{constant}$ --and in particular, the locus  $H = 0$  of the acoustic rays--are invariant under the  $F$ -transformations. But there is more: each integral curve of the  $H$ -equations retains this property--although it changes its individuality--under the  $F$ -transformations; and reciprocally, each integral curve of the  $F$ -equations is carried into another such curve by the  $H$ -transformations.

The proof of the latter "kinematical" property--which is a simple special case of the theorems of Lie groups--can be given by showing that the  $F$ -transformation preserves the property of a function's being a first integral of the  $H$ -equations. Thus if  $G$  is such an integral we have  $[G, H] = C$ , which, together with the given  $[F, H] = 0$  show, by

Jacobi's identity applied to the triplet  $F, G, H$ , that  $[[G, F], H] = 0$ , and hence that  $G + \sigma[G, F]$  is a first integral of the  $H$ -equations. Since this is  $G + \sigma(dG/d\sigma)$ --i.e., the result of applying the  $F$ -transformation to  $G$  (to first order quantities in  $\sigma$ )--the property of  $G$ 's remaining a first integral is proved to the first order: and to all orders, by the routine reasoning of differential equations. Since any integral curve of the  $H$ -equations can be represented as an intersection of  $2n$  independent first integrals, and vice versa, the kinetic property is thus proved. Similarly for the effect of the  $H$ -transformation on the  $F$ -integral curves.

Our first acoustical application of these theorems is to the laminar case. Here we have the two first integrals,  $H$  and  $F = p_r$ . We find the  $F$ -transformation by solving the corresponding canonical system, namely

$$(16.20) \quad \frac{dr}{d\sigma} = \frac{\partial F}{\partial p_r} = 1, \quad \frac{dz}{d\sigma} = \frac{\partial F}{\partial p_z} = 0, \quad \frac{dp_z}{d\sigma} = 0, \quad \frac{dp_r}{d\sigma} = -\frac{\partial F}{\partial z} = 0$$

The solution is:

$$(16.21) \quad r = r^0 + \sigma, \quad z = z^0, \quad p_r = p_r^0, \quad p_z = p_z^0.$$

This is a horizontal translation by the amount  $\sigma$ . If we look back at Figure 14.I, it becomes evident that such a translation leaves the  $p_r = k$  cylinders invariant, and carries the family of rays winding about it into the same family: every ray on the cylinder is congruent to every other, and the smallest positive increment of  $\sigma$  which carries a given ray into itself--which we have called  $R = R(k)$ --is their common period.

All this simply repeats what was shown in Appendix B and used throughout our treatment of the laminar case. (valid even when  $H = h \neq 0$ )

Our approach to the secularly non-laminar case is to operate as we have just done, but setting  $F = P_r(r, z, p_z)$  instead of  $F = p_r$ , since it is the former and not the latter which is a first integral of the ray equations. First, however, we must gain a more quantitative appreciation of this function. This is accomplished by solving the canonical ray equations (16.1) by a method of successive approximations based on the smallness of  $\epsilon$ .

We are not at this point setting  $H = 0$ ; we are however writing  $V = U + \epsilon W$ , where  $U = U(z)$  and  $W = W(r, z)$ , as explained earlier. We shall write  $H = H_0 + \epsilon W$ ,  $H_0$  being the laminar Hamiltonian. Equations (16.1) become

$$\frac{dr}{d\tau} = [r, H] = [r, H_0] + \epsilon [r, W]$$

$$\frac{dp_r}{d\tau} = [p_r, H] = [p_r, H_0] + \epsilon [p_r, W] = \epsilon [p_r, W]$$

together with the two further equations for  $z$  and  $p_z$ . The special form of the second equation results from the fact that  $[p_r, H_0] = 0$ : it will allow us to infer the closeness of  $p_r$  to its initial value  $p_r^0$ --and hence to  $p_r$ . It is explained early in Appendix C that the assumption that  $c$ , and hence  $n$ ,  $V$ , etc., are adequately represented

as analytic functions of the coordinates of position. With this assumption (and their linearity in  $\epsilon$ ), the theorems cited earlier (the G. D. Birkhoff reference) apply, and we can assume that all solutions are analytic, in all their arguments, and can in particular be expanded in convergent power series in  $\epsilon$ . Thus we can write in particular

$$(16.22) \quad r = r^0 + \epsilon r' + \epsilon^2 r'' + \dots, \quad p_r = p_r^0 + \epsilon p_r' + \epsilon^2 p_r'' + \dots$$

When these are inserted in the above equations and like powers of  $\epsilon$  equated, we obtain the sequence of equations defining the successive approximations

$$\begin{aligned} \frac{dr^0}{d\tau} - [r^0, H_0] &= 0 & \frac{dp_r^0}{d\tau} &= 0 \\ (16.23) \quad \frac{dr'}{d\tau} - [r', H_0] &= [r^0, W] & \frac{dp_r'}{d\tau} - [p_r', H_0] &= [p_r^0, W] = 0 \\ \frac{dr''}{d\tau} - [r'', H_0] &= [r', W] & \frac{dp_r''}{d\tau} - [p_r'', H_0] &= [p_r', W] \\ &\dots & &\dots \end{aligned}$$

together with another pair of columns for  $z$  and  $p_z$ . They are solved by recurrence, by a schema that can easily be developed as needed. For the present it is sufficient to observe that they show, on the basis of the definition of  $P_r(r, z, p_z)$  given earlier, that this function is of the form  $p_r + \epsilon \bar{P}_r(\epsilon, r, z, p_z)$ ; cf. (16.9), (16.11) and (16.22), etc.

Therefore, on setting  $F = P_r = p_r + \epsilon \bar{P}_r$  in (16.19), we obtain instead of (16.20) the following equations:

$$\frac{dr}{d\sigma} = 1 + \epsilon \frac{\partial}{\partial p_r} \bar{P}_r, \quad \frac{dz}{d\sigma} = \epsilon \frac{\partial}{\partial p_z} \bar{P}_r, \quad (16.24)$$

$$\frac{dp_r}{d\sigma} = -\epsilon \frac{\partial}{\partial r} \bar{P}_r, \quad \frac{dp_z}{d\sigma} = -\epsilon \frac{\partial}{\partial z} \bar{P}_r,$$

This--or its equivalent (16.19)--is our second Hamiltonian system.

These equations are again of the general type studied in Section 9; they define one and only one integral curve through each point  $(r, z, p_r, p_z)$  of phase space, given analytically, e.g., by expressing these quantities as functions of their initial values (indicated by zero superscripts) and the parameter  $\sigma$ . But since this variable does not occur explicitly in the coefficients in (16.24)--any more than  $\tau$  did in (16.1)--we may also use  $r$  as the independent variable,  $\sigma$  being eliminated from the equations just as  $\tau$  was in the earlier case. But the integrand of their fundamental sliding relative integral invariant is, by the general theory of Appendix E,

$$(16.25) \quad \omega^*(\delta) = p_r \delta r + p_z \delta z - F \delta \sigma, \quad F = P_r$$

Also, the Hamiltonian  $F$ , not containing  $\sigma$  explicitly, is a first integral of this system.

As noted before, our choice of  $F = P_r$  in (16.19) = (16.24) is such as to guarantee that (16.24) and (16.1) share the two first integrals  $H$  and  $F$ ; and further, that every  $F$ -transformation (determined by



a given value of  $\sigma$ ) shall carry any integral curve of (16.1) into an integral curve of the same system; and vice versa, any H-transformation (specified by  $\tau$ ) shall conserve the system of integral curves of (16.24). These facts, combined with the geometrically "simple" nature of the integral surfaces  $H = \text{constant}$  and  $F = \text{constant}$  allow the important relationships to be exhibited graphically, as shown in Figure 16.II.

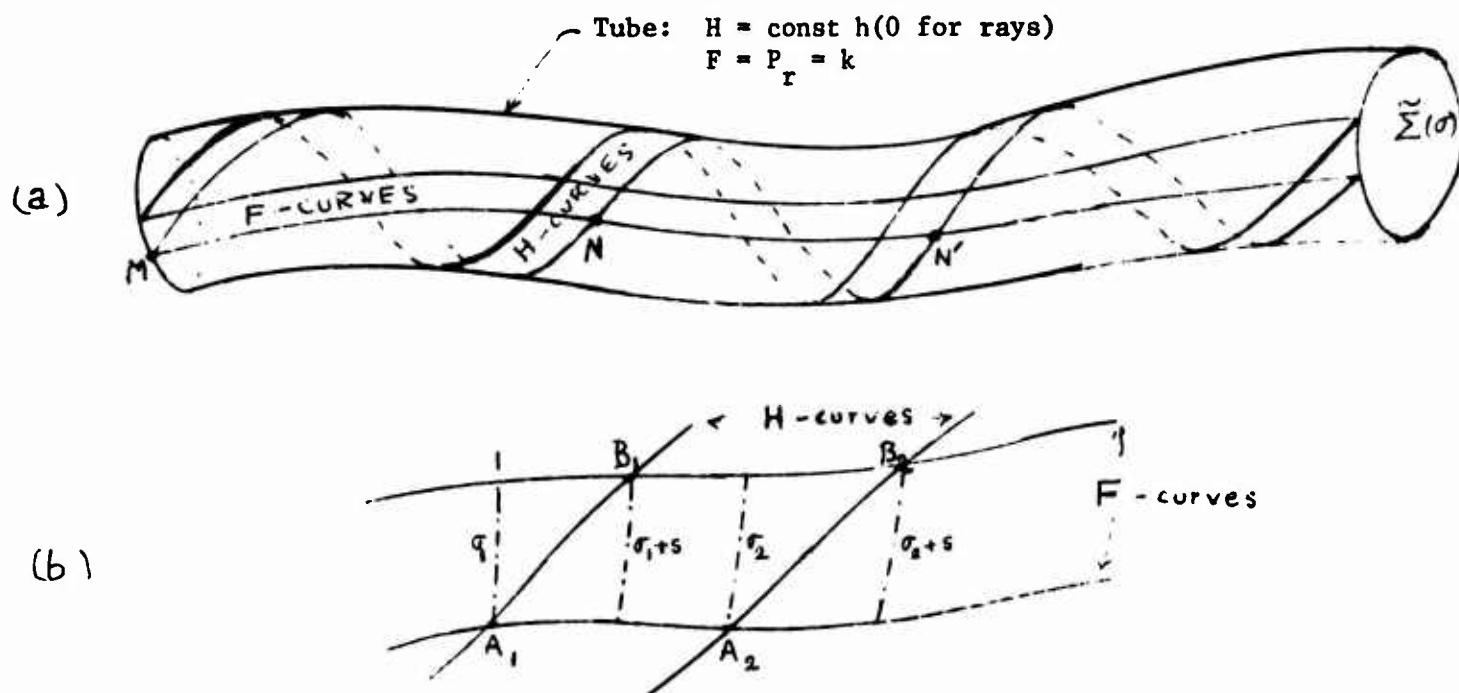


FIGURE 16.II: TUBE (a); GROUP PROPERTY (b).

Note that since (16.24) differs by quantities of the first order in  $\epsilon$  from (16.20)--the integral curves (16.21) being horizontal straight lines--we can suppose that those of (16.24) will be only moderately deformed versions of such lines, and hence will run along the "tubes"  $F = P_r = k$ , without winding around them, as do the integral curves (or rays, when  $H = 0$ ) of (16.1). As a matter of fact, regarding the slopes

of their projections on the  $rz$ -plane,  $\tan \alpha' = dz/dr$  calculated from (16.24) has a factor of  $\epsilon$ , while  $\tan \alpha = dz/dr$  calculated from (16.1) has no such factor: the former are in general much less steep than the latter. Finally, the continuous change in the two classes of integral curves with the continuous increase of  $\epsilon$  from 0 will deform the curves but not change their topological relations with the tubes or to one another--until  $\epsilon$  reaches some sharply determined critical value which, as part of our definition of the "secularity" of the non-laminarity, does not occur before  $\epsilon > 1$ .

These geometrical relations (or the equivalent analysis, given later) have two consequences: first, measuring horizontal shifts by increases in the parameter  $\sigma$  in (16.24) rather than in terms of  $r$ , an exact period  $\Pi$  can be defined for all the rays (the H-curves) on each tube  $F = k$ . For suppose that the unique H-curve and F-curve through the arbitrarily chosen point M on the tube  $F = k$  have their next intersection at N; and that  $\Pi$  is the increase in  $\sigma$  in the motion along the F-curve carrying M to N. Then the same increase will carry the arc  $\overline{MN}$  of the H-curve into another arc of H-curve,  $\overline{NN'}$ , which (by the uniqueness of these curves through any given point) will coincide with the whole H-curve through M; hence  $\Pi$  is the period of this particular H-curve. The fact that this same increase in  $\sigma$  brings every other H-curve on our tube into coincidence with itself is evident on the basis of the geometry and the other two properties just used. Of course this common period  $\Pi$  depends on the tube, i.e., the values of the constants in its equations ( $F = k$ ,  $H = h$ ). In the case of the rays ( $H = 0$ ) the tube and the corresponding period are fully characterized by the single constant  $k$ ; we write  $\Pi = \Pi(k)$ .

The second consequence is still more important: it is that we can adapt the construction of the surface of section  $\Sigma(r)$  to the present non-laminar case in such a way that the transformations induced in it by the rays will regain the group property.<sup>\*</sup> This shows that the theorems of the classical ergodic theory of mixing can be applied with the same authority to the case of secular non-laminarity as was done in Section 14 in the laminar case: the intuitive extension sketched earlier in the present Section is validated.

The modified definition of the surface of section and the proof of the group property may be indicated by geometrical constructions based on Figure 16.II; its analytic paraphrasing will follow the geometric. First, we take the initial surface of section  $\tilde{\Sigma}(r^0) = \Sigma(r^0)$  as before. Second, we increase the parameter  $\sigma$  from its initial value ( $\sigma^0 = r^0$ ) by an arbitrarily fixed amount, and denote by  $\tilde{\Sigma}(\sigma)$  the locus of points in phase space into which the F-transformation defined by (16.24) carries the points on  $\Sigma(r_0)$ . Third, we "identify"--regard as congruent--any point  $A_1$  on any  $\tilde{\Sigma}(\sigma_1)$  with the unique point  $A_2$  on any  $\tilde{\Sigma}(\sigma_2)$  whenever  $A_1$  and  $A_2$  lie on the same integral curve of (16.24) (the same F-curve). Fourth, we define the ray-induced transformation of points of  $\tilde{\Sigma}(\sigma_1)$  into those of  $\tilde{\Sigma}(\sigma_2)$  by the property that they are the intersections with these surfaces of the same H-curve (ray). The fact that this construction of images is self-consistent, and that it has the group property is based on the relations illustrated in Figure 16.II(b):

---

\* More exactly, the "group germ" property: a piece of a group, since the number of iterations is limited by permissable range.

$A_1$  and  $A_2$ , lying in  $\Sigma(\sigma_1)$  and  $\Sigma(\sigma_2)$  and on the same F-curve are, as we stated, congruent points. For any given  $s$ , let  $B_1$  be the image in  $\Sigma(\sigma_1 + s)$  of  $A_1$ --i.e.,  $B_1$  is the intersection with  $\Sigma(\sigma_1 + s)$  of the ray through  $A_1$ . Let  $B_2$  be the intersection with  $\Sigma(\sigma_2 + s)$  of the F-curve through  $B_1$ :  $B_2$  and  $B_1$  are congruent points in  $\Sigma(\sigma_2 + s)$  and  $\Sigma(\sigma_1 + s)$ . Now during the course of the F-transformation, as  $\sigma$  increases from  $\sigma_1$  to  $\sigma_2$ , the H-curve arc  $\overset{\frown}{A_1 B_1}$  through  $A_1$  is carried into an H-curve arc through  $A_2$ ; by the uniqueness of such arcs through given points, this must be precisely the arc  $\overset{\frown}{A_2 B_2}$ . Hence the identification of points on different surfaces  $\Sigma(\sigma)$  by the above rule of congruence is compatible with the definition of the ray-induced transformation on this surface of section.

The group property of the transformation induced in  $\Sigma(\sigma)$  by the H-curves is an immediate consequence of the above construction, combined with the group property of the H-transformations in phase space, noted earlier as the consequence of H not containing  $\tau$  (so that any solution of the H-equations retains this property--and gives the same integral curve--when  $\tau$  is increased by any constant  $t$ : the "motion" of points depends therefore only on this increase  $t$ , and not on the original "epoch"  $\tau$ ). Suppose as before that  $B_1$  is the image of  $A_1$  under the H-transformation when  $\tau$  is increased by  $t$ . Then by the considerations of the last paragraph,  $B_2$  is the image of  $A_2$  when the epoch (e.g.,  $\tau'$ ) is increased by the same amount  $t$ . The preceding paragraph showed that any class of congruent points (i.e., lying on the same F-curve) will go into a class of congruent points under an H-transformation: here we have shown that the particular class the former goes into is fully

determined by the increase  $t$  in the parameter  $\tau$ —whatever the value of the latter. Thus successive transformations of congruent classes, first by an increase  $t$ , then by  $t'$ , will have the same result as the single transformation produced by the increase  $t + t'$ : this is the group property.

It will be noted that when in (16.24)  $\epsilon = 0$ , (16.24) coincides with (16.20), and all our constructions reduce to those of Section 14 for the laminar case.

In the geometric representation of Figure 16.II, the surface of section  $\Sigma(\sigma)$  is not a plane, but may be described as a slightly deformed plane region: the  $\Sigma(r)$  for the  $r$  chosen as the solution of (16.24) calculated for a fixed value of  $\sigma$  in  $\Sigma(\sigma)$ . The transformation induced by the rays in this surface of section is conservative, since  $\omega^*(\delta)$  (which on  $\Sigma(\sigma)$  has  $\delta\sigma = 0$ ) when integrated about the contour of any piece  $\Delta\Sigma$  is a sliding invariant: we might call it an "invariant mass"; but in the above representation it is not an area, although it reduces to an area as  $\epsilon \rightarrow 0$  and the surface of section reduces to  $\Sigma(r)$ . In another geometrical representation, in the system of coordinates introduced in the canonical transformation given below, the surface of section and its invariant mass will regain the form of planes and areas.

To sum up the results achieved by the introduction of the second Hamiltonian based on  $F = P_r$ : they prove that in the secularly non-laminar case those results of Section 14 that led, in the laminar case, to (14.10)' and hence to our equations (16.13) through (16.16) apply. On the other hand, to pass from this to the final transmission loss

formula (14.20), we are still obliged to fall back on the earlier intuitive approach to the second part of the definition of secular non-laminarity.

A canonical change of variables. We can go further and give the preceding somewhat intuitively geometrical treatment a firm analytical underpinning. This is done by making a canonical change of variables in our phase space--i.e., one which preserves the Hamiltonian properties. The new variables are suggested by applying the general method, so basic to all the work in the present section, of looking for "geometrically simple" first integrals of the "unperturbed" (laminar) system, and using the fact that (by the theorems in the reference to G. D. Birkhoff cited above) this geometrical simplicity is retained in the "perturbed" system ( $\epsilon > 0$ ).

In the case of (16.1), the only geometrically simple first integrals even when  $\epsilon = 0$  were  $H$  and  $p_r$ --our whole treatment has been based on the cylindrical locus, tubular when  $\epsilon > 0$ , obtained by setting  $H$  and  $P_r$  equal to constants.

In the case of (16.24), on the other hand, every one of the first integrals is geometrically simple when  $\epsilon = 0$ , as we see by reference to (16.20) and to its solutions given there. Hence the perturbed ( $\epsilon > 0$ ) integrals, i.e., those of (16.24), are geometrically simple. Our change of variable may be described as a moderate deformation of phase

space through which the somewhat bent integral surfaces are "flattened out"--and which at the same time preserves the Hamiltonian form of the new system (16.1). Since integral curves are intersections of integral surfaces, this bending will bring the F-curves into horizontal straight lines--as were those of (16.20) when  $\epsilon = 0$ .

The analytical technique is to make a canonical change of variables. The methods, insofar as they involve local properties of the functions, are old and familiar. They are set forth in Whittaker (l.c., above), and have been applied in a restricted form in the old quantum theory (see, e.g., the 4th and 7th mathematical appendices of A. Sommerfeld, "Atombau und Spektrallinien", Braunschweig, 1922). But since long range propagation requires going beyond local properties, a special form has to be given to the present application of these methods.

To give perspective and gain typographical simplicity, we shall consider the general local problem first: It is desired to change from  $n$  position coordinates  $q_i$  and conjugate momenta  $p_i$  to new  $\bar{q}_i, \bar{p}_i$ , in such a way as to preserve the Hamiltonian form of any given system such as (16.1) or (16.14). In general, this requires that  $\sum p_i \delta q_i$  differ from  $\sum \bar{p}_i \delta \bar{q}_i$  by an exact differential of a function of any  $2n$  independent variables in the set of  $4n$ . There is a practical advantage (pointed out by Sommerfeld, l.c.) and a theoretical advantage (the group property, pointed out by G. D. Birkhoff, l.c., II.11) in writing the above condition in a slightly modified form. Applying  $\delta(\bar{p}_i \bar{q}_i) = \bar{p}_i \delta \bar{q}_i + \bar{q}_i \delta \bar{p}_i$ , etc., we can write the condition for a canonical (sometimes called a

"contact") transformation as follows:

$$(16.26) \quad \sum_1 p_1 \delta q_1 + \sum_1 \bar{q}_1 \delta \bar{p}_1 = \delta G(q_1, \dots, q_n; \bar{p}_1, \dots, \bar{p}_n)$$

which, when  $(q, \bar{p})$  are  $2n$  independent variables, is equivalent to

$$(16.27) \quad p_1 = \frac{\partial G}{\partial q_1}, \quad \bar{q}_1 = \frac{\partial G}{\partial \bar{p}_1}.$$

The property of  $(q, \bar{p})$  being independent, and the possibility of solving (16.27) for  $(q, p)$  in terms of  $(\bar{q}, \bar{p})$  are--at the local level--equivalent to one another and to the non-vanishing of a certain functional determinant (Jacobian). The latter reduces to the determinant  $\Delta = |\partial^2 G / \partial q_i \partial \bar{p}_j|$ .

When we take  $G = \sum_1 q_1 \bar{p}_1$ , the transformation defined by (16.27) obviously reduces to the identity:  $\bar{p}_1 = p_1$ ,  $\bar{q}_1 = q_1$ . Our "slight deformation" of the phase space of  $(r, z, p_r, p_z)$  will make use of a function  $G_\epsilon(r, z, \bar{p}_r, \bar{p}_z)$  which reduces to  $r\bar{p}_r + z\bar{p}_z$  when  $\epsilon = 0$  (no deformation in the laminar case!). We also wish to have no deformation "initially", i.e., when  $r = r^0$ , so that

$$(16.28) \quad G_\epsilon(r^0, z, \bar{p}_r, \bar{p}_z) = r^0 \bar{p}_r + z \bar{p}_z \quad (\epsilon \geq 0)$$

The final requirement is that (16.24) should reduce to (16.20) in the barred coordinates: this is the "straightening out" of its integral curves. Inasmuch as in the un-barred coordinates  $F(q, p) = k$  (a constant) along each integral curve,  $G$  must, in view of (16.27), satisfy the Hamilton-Jacobi partial differential equation



$F(q, \partial G/\partial p) = k$ ; i.e.

$$(16.29) \quad F(r, z, \frac{\partial G}{\partial r}, \frac{\partial G}{\partial z}) = k$$

Now since  $F = P_r = p_r + \epsilon \bar{p}_r$ , as shown above,  $\partial F/\partial p_r$  will not vanish for the moderate values of  $\epsilon$ . Hence the solution of (16.29) with the boundary condition (16.28) is a "Cauchy problem" having a unique solution--and one which, by the geometrically simple form of its characteristics (the integral curves of (16.24)) has a unique solution, valid throughout the range interval  $(r^0, r^1)$  of importance in the present section. For a proof, see e.g., Osgood, l.c., (XIV, 24, 25).

When  $\epsilon = 0$ , (16.29) reduces to  $\partial G/\partial r = k$ , whose solution subject to (16.28) is  $r\bar{p}_r + z\bar{p}_z$  (since we are defining  $k$  as  $\bar{p}_r$ ). When  $\epsilon > 0$ ,  $G$  can be expressed as a convergent series in powers of  $\epsilon$ , starting with the above expression: the presence of the constants  $(\bar{p}_r, \bar{p}_z)$  in this and in condition (16.28) shows that  $G_\epsilon$  depends on them: it is a function of  $(r, z, \bar{p}_r, \bar{p}_z)$ . Furthermore,  $\Delta = 1$  when  $\epsilon = 0$  and differs from unity by terms of the first order in  $\epsilon$  when  $\epsilon > 0$ , so we can assume  $\Delta > 0$  for the slight change as  $\epsilon = 1$  ("secularity"). Finally, in the new (dashed) variables,  $F = \bar{p}_r = k$ , so that equations (16.24) reduce to (16.20) in these variables: the straightening out has been accomplished in the large--not just locally (as in the text-books<sup>\*</sup>).

Turning to (16.1), we must re-write it in terms of the dashed letters, using the new Hamiltonian

---

\* H. Poincaré<sup>5</sup> being an honorable exception.

$$\bar{H}(\bar{r}, \bar{z}, \bar{p}_r, \bar{p}_z) = H(r, z, p_r, p_z),$$

obtained by simple substitutions of the undashed variables on the right by their functional expressions in terms of the new variables, defined by (16.27). [ $\bar{H}$  is a different function of its four variables than is  $H$  of its variables: their values are equal at points corresponding by the transformation (16.27)]. From the fact that  $[F, H] = 0$  we infer that in the new variables  $[\bar{F}, \bar{H}] = 0$  also, this being a formal property of Poisson brackets under canonical transformations in phase space. Therefore, since  $\bar{F} = \bar{p}_r$ , we see that  $\partial \bar{H} / \partial \bar{r} = 0$ : our change of variables has restored laminarity--together with all the properties and constructions of the previous sections that were based on that assumption--with one exception, to be noted below.

The fact that the two equations  $[F, H] = 0$  and  $[\bar{F}, \bar{H}] = 0$  are equivalent can be made evident geometrically, since they each express the same fact--that the  $F$ -transformation carries arcs of  $H$ -curves into such curves, and vice versa: a relationship independent of the choice of coordinates.

The rays correspond with the subspace  $\mathfrak{S}_3$  for which  $\bar{H} = 0$  and in which the integrand of the sliding relative integral invariant is

$$\begin{aligned}\bar{\omega}(\delta) &= \bar{p}_r \delta \bar{r} + \bar{p}_z \delta \bar{z} - \bar{H} \delta \tau \\ &= \bar{p}_r \delta \bar{r} + \bar{p}_z \delta \bar{z}\end{aligned}$$

from which we derive, by the generalized Stokes' theorem, that  $\delta \bar{p}_r \wedge \delta \bar{r} + \delta \bar{p}_z \wedge \delta \bar{r}$  is an absolute sliding integral invariant, reducing to the element of area on any surface  $\bar{r} = \text{constant}$ . Since in the new variables the F-curves are the horizontal lines, with  $\bar{r} = \sigma$ , it is evident that the modified surface of section  $\tilde{\Sigma}(\sigma)$  coincides with  $\bar{\Sigma}(\bar{r})$ : constructed exactly as was  $\Sigma(r)$  in the laminar case. Finally, the power flux density  $\tilde{f}$  across  $\bar{\Sigma}(\bar{r})$  is defined precisely as in Section 14--simply replacing un-dashed by dashed variables. By this method, the definition of  $\tilde{f}$  and its substitution for  $f$ , together with the formulas (16.13) through (16.16) are validated--without the extra appeal to the intuition that was made in their derivations earlier in this Section.

The exception noted above has to do with the figure into which our "slight deformation" changes the horizontal line  $z = \zeta$  or band  $\zeta' < z < \zeta''$  at the receiver, where  $r = r^1$ ; cf. Figure 16.I. Evidently this shares the deformation and becomes bent--slightly but irregularly. We might re-interpret Figure 16.I(b) by regarding the symmetrical curve (i)(i)(i) . . . as the locus  $\bar{p}_r = k$  in the space in which the dashed variables are the coordinates, so that it is the result of applying our deformation to (j)(j)(j) . . . , and since canonical transformations preserve  $\int \omega$  and  $\int \tilde{\omega}$ , the area of (i)(i)(i) . . . will be equal to that of (j)(j)(j) . . . --as it was by construction in its earlier interpretation. On the other hand this deformation changes the horizontal receiver depth line  $z = \zeta$  and band  $\zeta' < z < \zeta''$  into slightly but irregularly bent figures, that are not horizontal ( $\bar{z} \neq \text{constant}$ ). To establish the final transmission loss formulas, we are again obliged to use an

intuitive interpretation of the "secularity" of the departure from laminarity. We can, as before, extend our interval by the additional one  $(r^1, r^2)$  throughout which the acoustic profile no longer changes; and then replace the set of images of  $z = \zeta$ , etc., as  $r$  increases over this interval by a sort of average or equivalent horizontal line  $\bar{z} = \bar{\zeta}$ , etc.

It is desirable, however, to give a "cleaner" formulation in precise mathematical terms, from which the desired formula can be obtained as a logical deduction. This will be done in our third and final treatment of the transmission loss problem in the case of secular non-laminarity. We shall make use of a variety of known theorems and traditional methods, but bring them to bear on our problem in a manner that is new. It has the advantage of generalizability to xyz-space.

The Method of Simultaneous Hamilton-Jacobi Equations. Our starting-point is the material set forth up to our intuitive approach, and also the early part of our second more rigorous treatment in which the Poisson brackets are introduced, formulas (16.17) through (16.25) are developed, and Figure 16.II established. Our method will be to make a canonical transformation of the type (16.26) and (16.27); but far more drastic in its effect: it will not only straighten out the surfaces  $F = k$  into horizontal cylinders, but it will deform the H-curves that wind around them into actual helices--which become straight lines when the cylinders, thought of as made out of paper, are cut along an element and flattened out into planes.

Analytically, this takes the form of the introduction by the canonical transformation of "angle" and "action" variables. The latter are constant along each integral curve (they are first integrals of the differential equations (16.1)) the former increase at a constant rate, depending on the values of the former. Finally, the original coordinates (our present  $r, z, p_r, p_z$ ) are functions of the angle and action variables, either proportional to, or periodic functions of, the angle variables. This will allow, among other things, a perfectly precise statement of the second part of the definition of "secular non-laminarity".

The introduction of angle and action variables by a canonical transformation has long been familiar in celestial mechanics, where it is applied to perturbation theory. It had a temporary revival in the old (pre 1924) quantum theory (see Sommerfeld, l.c.). But in all such cases the potential in the Hamiltonian was a sum of functions of single position coordinates (when chosen in a proper way), and the Hamilton-Jacobi equation could be solved explicitly by the method of "separation of variables"; therefore the answer to the whole problem could be given in terms of integrals of known functions. This method was applied to the unperturbed problem, the original coordinates were expanded in multiple trigonometric series in the angle variables, and thus a convenient starting point was obtained for a method of successive series approximations--complicated, but well fitted to modern computers. Some of the same ideas were used in the formulation of the quantum conditions in the old theory and, in particular, to show their "adiabatic invariance"

--a property of great interest at that time.

The present method makes the canonical transformation to angle and action variables in the actual ("perturbed") case:  $V = V(r, z)$ , making no assumption whatsoever concerning the separability of variables.

Its basis is to construct a line integral which is a solution of the pair of simultaneous Hamilton-Jacobi equations

$$\begin{aligned} (16.30) \quad & H(r, z, \partial G/\partial r, \partial G/\partial z) = h \\ & F(r, z, \partial G/\partial r, \partial G/\partial z) = k \end{aligned}$$

The fact that they are consistent--that they have a common solution  $G = G(r, z, h, k)$ --is a consequence of the fact that for our present  $H$  and  $F$ ,  $[H, F] = 0$ ; indeed this is necessary as well as sufficient for the consistency of (16.30)\*. Unfortunately the standard theory of these equations is either confined to special simple cases, or, when general, is purely local. When its results are extended to less restricted intervals of the variables, as is required in long range propagation, the functions develop singularities, become multiple-valued, or even cease to exist: a difficulty already noted in Sections 1 and 2, with the solutions  $G = S$  of the first equation in (16.30) with  $h = 0$ --the "eikonal".

---

\* Cf. any treatment of simultaneous non-linear first order partial differential equations; e.g., E. Goursat "Leçons sur l'Intégration des Equations aux Dérivées Partielles du Premier Ordre" (A. Hermann, Paris, 1891).

In the following treatment the solution of (16.30) that we construct as a line integral will be shown to have the properties needed for the canonical transformation to angle and action variables. The variables  $r, z, p_r, p_z$  are subjected to the restrictions

$$(16.31) \quad 0 \leq r^0 \leq r \leq r^1; a \leq z \leq b; h' < H < h''; p_r \leq n(r, b).$$

Here  $(h', h'')$  is an arbitrarily small but fixed interval containing  $H = h = 0$  at its interior. Since  $r^1 - r^0$  is the range of propagation, we are evidently not confining ourselves to "local" mathematics. As a consequence of the first part of our definition of secular non-laminarity, under the restrictions (16.31) (which imply restrictions on  $k = F(r, z, p_r, p_z)$ ) whenever (correspondingly restricted) values of  $h$  and  $k$  are given, the locus  $H = h, F = k$ , has the form of the tube shown in Figure 16.II(a). This is reproduced in Figure 16.III, together with certain construction lines.

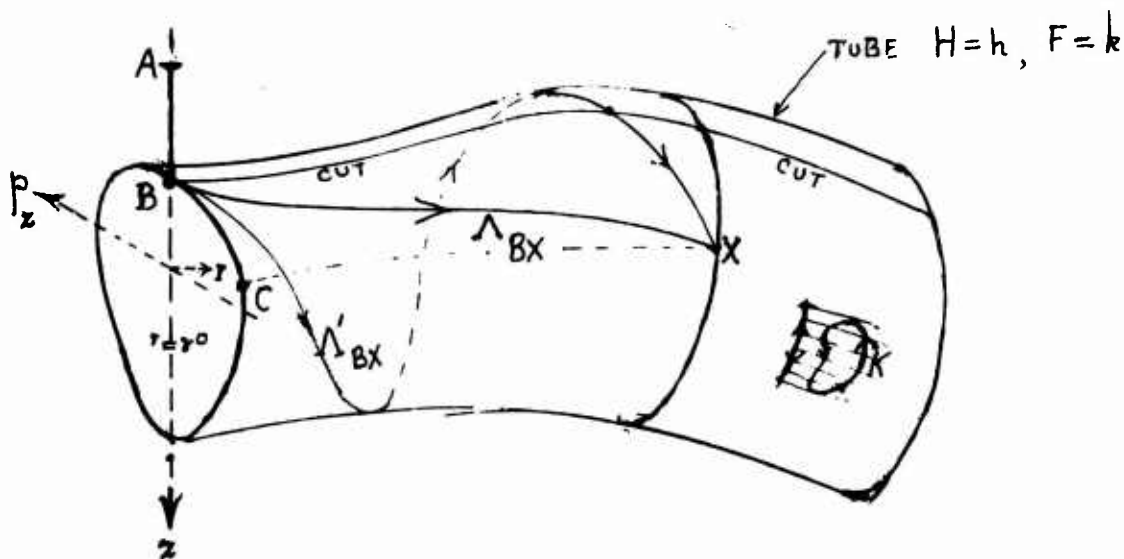


FIGURE 16.III

Let values of  $(r, z, h, k)$  satisfying (16.31) be given. The latter two determine the tube shown in Figure 16.III, while  $r$  determines the closed curve of its intersection with the vertical plane of given  $r$ . As for  $z$ , there may be no point, one point, or two points on this closed curve having precisely the depth  $z$ . Clearly, if we are to use  $(h, k)$  as "momenta" instead of  $p_r, p_z$ , we need a more adequate position coordinate than  $z$ : this will introduce itself automatically in the analysis that follows.

Instead of starting with  $(r, z, h, k)$  we shall start with  $(X, h, k)$ , where  $X$  is a point chosen arbitrarily on the tube of Figure 16.III, its range  $r$  therefore lying in the interval  $(r^0, r^1)$ . We shall define a function  $G(X, h, k)$ --actually infinitely multiple-valued, of simple type--as the line integral of  $p_r \delta r + p_z \delta z$  along the path  $\Lambda = \Lambda(X, h, k)$  shown in the figure: the path starts at  $A: (r^0, a, p_r, 0)$  on the surface ( $p_r$  being the non-negative solution of  $H = h$ ). It follows the vertical line segment  $AB$  to the point  $B$  where it first meets the tube determined by  $(h, k)$ . Then it stays on the tube, taking any path  $\Lambda_{BX}$  from  $B$  to  $X$ .

The arbitrariness of the path  $\Lambda_{BX}$  is what leads to the multiple-valuedness of the line integral; but this is of a restricted type, as can be seen as follows: let  $K$  be any closed curve on the tube that can be continuously shrunk to a point ("homotopic to zero"); then by sliding along the integral curves of the  $H$ -equation (which make up the tube), path  $K$  can be reduced to a line between two points, traced first in one



direction, then in the opposite one. Hence  $\int \omega^*$  of (16.25) around  $K$  is zero, because of its being a sliding (relative) integral invariant. Since the term  $F\delta\sigma$  in  $\omega$  is equal to  $k\delta\sigma = \delta(k\sigma)$  on the tube (and  $\sigma$  is single-valued on it), its contribution to the integral around  $C$  is zero. Therefore  $\int_K p_r \delta r + p_z \delta z = 0$  (this equation could also have been derived from the  $\omega$  of (16.17) using  $H$  instead of  $F$ ).

From the fact thus established it follows that our integral along  $BX$  will have the same value for all paths on the tube that join  $B$  to  $X$ , provided one path can be moved continuously on the tube into the other; but not in general for two paths that cannot be so moved (non-"homotopic"). Let  $J = J(h, k)$  be the integral of  $p_r \delta r + p_z \delta z$  (or equivalently of  $\omega$  or  $\omega^*$ ) around a single loop of the tube, taken in the  $+p_z \rightarrow +z$  clockwise direction (negatively, with respect to the coordinate axes  $r, z, p_z$  in  $\mathcal{F}_3$ ). As we know, this sliding integral invariant is equal to the common area of intersection of the tube with any plane  $r = \text{constant}$  (cf. Sections 13 and 14). Evidently any path  $\Lambda_{BX}$  can be reduced to any other, combined with an integral number of loops of the above type, described either clockwise or counter-clockwise. Therefore the value of  $G(X, h, k)$  is determined up to a positive or negative integral multiple of  $J(h, k)$ .

We also note that the part of the integral along the vertical segment of the path,  $AB$ , is zero, since along it  $\delta r = 0$  and  $p_z = 0$  and thus integrand vanishes.

The usual methods of the advanced calculus show easily that

$$(16.32) \quad \frac{\partial}{\partial r} G(X, h, k) = p_r, \quad \frac{\partial}{\partial z} G(X, h, k) = p_z;$$

but this has to be interpreted: the point  $X$  determines the values  $(r, z)$ , and a sufficiently small increment  $\Delta r$  or  $\Delta z$  in either determines a new position  $X'$  of  $X$ ; when  $G(X', h, k) - G(X, h, k)$  is divided by the corresponding increment in the coordinate and the limit taken as the latter approaches zero, (16.32) is an easy result. If the initial  $X$  were at a highest or lowest point, the  $\Delta z$  would have to be negative or positive, respectively; there would be ambiguity in the position of  $X'$ , but not in  $\partial G/\partial z$ , which would be zero (cases in which the tube touches the water surface cause no material difficulty; actually they can be treated by the mathematical device of the "method of images").

It is necessary to calculate the partial derivatives of  $G(X, h, k)$  with respect to  $h$  and  $k$ ; i.e., holding  $X$  fixed. Since  $X$  was chosen on the tube after it had been determined by  $h$  and  $k$ , this requires explanation:  $X$  is to be thought of as specifying a definite position in the physical space of the medium of propagation, the  $rz$ -plane; to keep it fixed when (small) changes are made in  $h$  or  $k$ , resulting in changes in the shape and position of the tube, means that the  $(r, z)$  corresponding to  $X$  are not to change and therefore that it merely slides along a horizontal straight line in the  $(r, z, p_r, p_z)$  - phase space, along which the range  $r$  and depth  $z$  are constant.

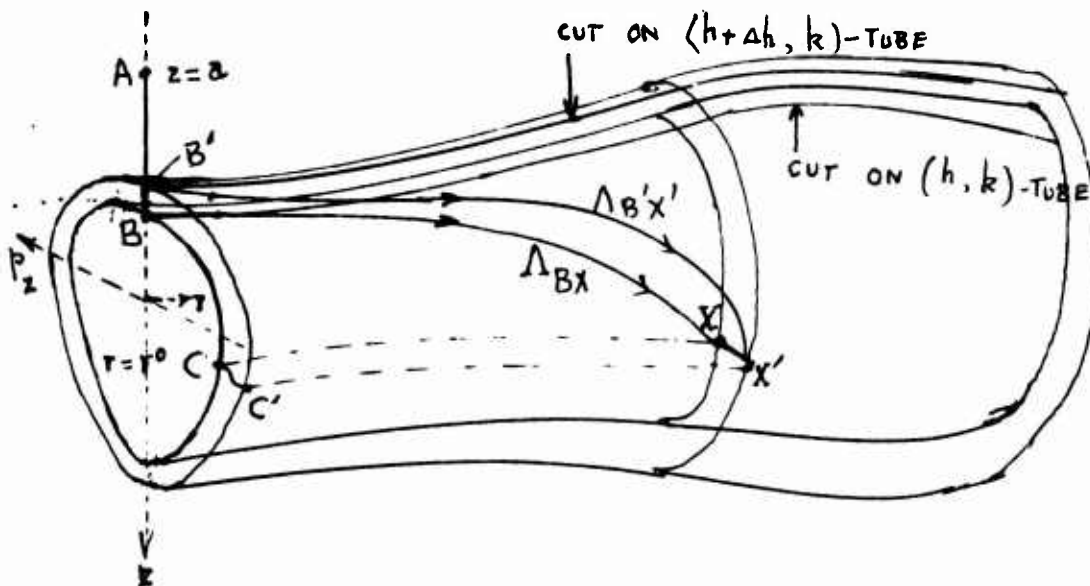


FIGURE 16.IV

First we shall calculate  $\partial G/\partial h$ . When  $h$  changes to  $h + \Delta h$  (remaining within the interval  $(h', h'')$ ), the  $(h, k)$  - tube changes to a  $(h + \Delta h, k)$  - tube,  $X$  sliding to a new position  $X'$  along the horizontal line described above. To find the effect of this change on the value of  $G(X, h, k)$ , we must remember that this function is multiple-valued, two values differing by integral multiples of  $J(h, k)$ , which will itself have a non-zero derivative with respect to  $h$ ; we shall separate our problem into two parts: finding the derivative of a particular branch of the function  $G_1(X, h, k)$  and then that of  $J(h, k)$ . Figure 16.IV shows the basis of the definition of what we shall take as the "principal branch"  $G_1$ , by standardizing the path  $\Lambda$ .

First we render our tubes simply connected by making a "cut" in all of them: through the highest point (least  $z$ ) where  $r = r^0$  and  $p_z = 0$  (the points  $B$  and  $B'$  on the two tubes shown) we pass the unique  $F$ -curve. This amounts to cutting along the surface generated by all the  $F$ -curves passing through the segment of the  $z$ -axis, from the point on the surface  $z = a$  to the central point  $z = \bar{z}$  of the duct (at which  $c$  is minimum,  $n$  maximum). We now restrict our paths  $\Lambda(BX)$ ,  $\Lambda(B'X')$ , etc. in the definition of  $G_1(X, h, k)$  by never allowing them to intersect the respective cut between their two extremities  $B, X$  or  $B', X'$ . Except for this restriction, these paths can be taken arbitrarily on their respective tubes: since every such path can be moved continuously into any other having the same extremities, because of the simple connectivity of the cut tube, the corresponding line integrals are all equal and therefore are functions of  $X$  (or  $X'$ ) only. Of course the values they approach as  $X$  approaches a point on the cut from either side differ by  $J(h, k)$ .

We now take the line integral of

$$(16.25) \quad \omega^* = p_r \delta r + p_z \delta z - F \delta \sigma$$

around the closed path  $AB'X'XBB'A$  in that order:  $B'X'$  and  $XB$  along the  $\Lambda$  curves restricted as above. Since throughout the region containing this path  $F = k$  (only  $H = h$  changes at present), the last term in  $\omega^*$ ,  $F \delta \sigma = k \delta \sigma = \delta(k\sigma)$  makes a zero contribution to the integral, the path being closed. On the line segment  $XX'$ ,  $r$  and  $z$  are constant, as we noted before; therefore  $\delta r = \delta z = 0$ . Therefore the line integral about

the closed path reduces to that of  $p_r \delta r + p_z \delta z$  along  $AB'X'$ , minus that along  $ABX$ : by definition of  $G_1$  this is

$$\Delta G_1 = G_1(X', h + \Delta h, k) - G_1(X, h, k)$$

To find its value we make use of the fact that  $\omega^*$  is the integrand of a sliding relative integral invariant of the F-equations (16.24): we can slide the path along the F-curves into the initial plane  $r = r^0$  without changing its value. If  $C$  and  $C'$  are the points into which  $X$  and  $X'$  are slid in this process, the path of integration becomes (leaving out  $AB'$  which makes a zero contribution)  $B'C'CBB'$ . But in this plane,  $\delta r = 0$ , so that the integrand is  $p_z \delta z$ : we know that its integral is the area in the  $zp_z$ -plane enclosed by the path. Hence  $\partial G / \partial h$  is the limit of the ratio of this area to  $\Delta h$  as  $\Delta h \rightarrow 0$ .

This limit is easy to find, since on the initial plane  $r = r^0$ ,  $F$ , which was defined as  $P_r$ , reduces to  $p_r$ , and we have  $p_r = k$ . The curves on which the arc  $BC$  lies has the equation

$$(16.33) \quad k^2 + p_z^2 = n_0^2 + 2h, \quad n_0 = n(r_0, z)$$

while that of  $B'C'$  is what this becomes when  $h$  is replaced by  $h + \Delta h$ .  $B'B$  is a straight line but  $CC'$  is merely a smooth curve cutting the two arcs. We note that when each of these short curves is replaced by any other smooth curves in the initial plane through  $B$  and  $C$ , approaching zero as  $\Delta h \rightarrow 0$ , the included area will only be altered by infinitesimals of higher order (i.e., a vanishing percentage as  $\Delta h \rightarrow 0$ ). Therefore we are able to replace them by horizontal line segments ( $z = \text{constant}$ )

through B and C: denote their depths by  $z_1$  and  $z_C$ . Then if  $p_z \geq 0$  at C, the area is

$$\int_{z_1}^{z_C} \sqrt{n_o^2 - k^2 + 2(h + \Delta h)} \, dz - \int_{z_1}^{z_C} \sqrt{n_o^2 - k^2 + 2h} \, dz$$

Dividing by  $\Delta h$  and letting  $\Delta h \rightarrow 0$  is equivalent to differentiating the second integral with respect to  $h$ . We obtain

$$(16.34) \quad \frac{\partial G_1}{\partial h} = \int_{z_1}^{z_C} \frac{dz}{\sqrt{n_o^2 - k^2 + 2h}}$$

When  $p_z < 0$  at C, the area is best written as the area of the whole ring between the two curves of the type (16.33), minus the area given of the rest of the ring. By the obvious ~~symmetry~~ symmetry in the  $z$ -axis, the area subtracted leads in the limit to a subtraction of the integral on the right in (16.34). On the other hand, the total area leads to the expression (16.36) derived below.

We now remove the cuts on the tubes, so that  $G$  becomes multiple-valued and may be written  $G_1 + NJ$ , where  $N = 0, \pm 1, \pm 2$ , etc. The derivative with respect to  $h$  will be given by that obtained above, to which is added  $N$  times that of  $J$ . The latter, being the integral of  $\bar{\omega}$  about a single loop around the tube, can be evaluated by sliding it into the plane  $r = r^0$ , where it becomes, as we know, the area enclosed by the curve (16.33):

$$(16.35) J(h, k) = \oint \sqrt{n_o^2 - k^2 + 2h} \cdot dz = 2 \int_{z_1}^{z_2} \sqrt{n_o^2 - k^2 + 2h} \cdot dz$$

Its h derivative is therefore

$$(16.36) \quad J_h(h, k) = 2 \int_{z_1}^{z_2} \frac{dz}{\sqrt{n_o^2 - k^2 + 2h}}$$

We obtain finally

$$(16.37) \quad \frac{\partial G}{\partial h} = N J_h(h, k) \pm \int_{z_1}^{z_c} \frac{dz}{\sqrt{n_o^2 - k^2 + 2h}}$$

The evaluation of  $\partial G / \partial k$  is carried out along the same lines.

The second tube, corresponding to the replacement of  $k$  by  $k + \Delta k$ , is inside the first; but the path  $AB'X'XBA$  is defined as before (Fig. 16.IV).

In the integral of  $\bar{\omega}$  around it, we obtain, in addition to the difference  $G_1(h, k + \Delta k) - G_1(h, k)$ , a non-vanishing contribution of  $F\delta\sigma$ . To calculate this, we note that on the part of the path that lies on the initial plane  $\delta\sigma = 0$ . There remains

$$\int_{\Lambda_{B'X'}} F\delta\sigma - \int_{\Lambda_{BX}} F\delta\sigma + \int_{X'X} F\delta\sigma$$

$$= (k + \Delta k) (\sigma_{X'} - \sigma^0) - k(\sigma_X - \sigma^0) + \bar{k}(\sigma_X - \sigma_{X'})$$

Where  $\sigma_X$  and  $\sigma_{X'}$  are the values of  $\sigma$  measured along the F-curves connecting  $X$  and  $X'$  to the initial plane, and  $\bar{k}$  is a number between  $k$  and

$k + \Delta k$ . The above expression reduces to

$$(k - \bar{k}) (\sigma_{X'} - \sigma_X) + \Delta k (\sigma_{X'} - \sigma^0)$$

On dividing by  $\Delta k$  and observing that  $(k - \bar{k})/\Delta k$  is bounded, while  $\sigma_{X'} \rightarrow \sigma_X$ , we see that the contribution is, finally,  $\sigma_X - \sigma^0$ .

Again we slide our closed path along the F-curves into the initial plane, obtaining a ring-shaped region whose boundaries lie on the curve (16.33), and a corresponding one with  $k$  replaced by  $k + \Delta k$ , together with the two short curves joining their pairs of extremities at  $z = z_1$  and  $z = z_c$ . The latter are, as before, modified to become horizontal lines of these depths. We have therefore the same difference of integrals as in the former case,  $k$  being different in the two integrands instead of  $h$ . Division by  $\Delta k$  and passage to the limit as this approaches zero gives, in the case that  $p_z \geq 0$  at  $z_c$ ,

$$(16.38) \quad \frac{\partial G_1}{\partial k} = \sigma - \sigma_0 - k \int_{z_1}^{z_c} \frac{dz}{\sqrt{n_0^2 - k^2 + 2h}}$$

The case in which  $p_z < 0$  at  $z_c$  is handled as before, replacing the line integral by the  $k$ -derivative of  $J$  minus the term containing the integral in (16.38). We may note that the appearance of the negative sign in this term corresponds to the fact that when  $\Delta k > 0$ , the contour integral when slid into the initial plane, is in counterclockwise sense, and so represents the negative of the enclosed area. This is because Snell curves and their tubes contract as  $k$  increases, as is shown by (16.33).



We easily find the k-derivative of J:

$$(16.39) \quad J_k(h, k) = -kJ_h(h, k).$$

Hence, finally, for the multiple-valued G,

$$(16.40) \quad \frac{\partial G}{\partial k} = \sigma - \sigma_0 - k \frac{\partial G}{\partial h},$$

the last term being given by (16.37).

We could now take  $G = G(r, z, h, k)$  in making a transformation such as (16.27), introducing the momenta  $(\bar{p}_1, \bar{p}_2) = (h, k)$ , and new position variables, given by the right-hand members in (16.37) and (16.40); but the results would be unnecessarily complicated, and, in fact, lack a certain valuable "intrinsic" quality. Rather than use the pair  $(h, k)$  we shall use the related pair,  $(j, k)$ , where  $j$  is the action variable, defined by the equation

$$(16.41) \quad j = J(h, k).$$

Since, for any given  $k$ ,  $J(h, k)$  is the area enclosed by the curve (16.33), and since as  $h$  increases across its interval  $(h', h'')$  this curve expands and so its area increases continuously, (16.41) determines  $h$  as a function of  $j$  and  $k$  (within the appropriate intervals) which we may write  $h = h(j, k)$ , and apply the rules of differentiation of implicit functions to calculating its derivatives. Thus

$$\frac{\partial h(j, k)}{\partial j} = 1/J_h(h, k),$$

$$\frac{\partial h(j, k)}{\partial k} = -J_k(h, k)/J_h(h, k) = k,$$

the latter by virtue of (16.39).

We now express  $G$  in terms of  $j, k$ , writing

$$\Gamma(X, j, k) = G(X, h(j, k), k)$$

It follows at once from (16.37) and (16.40) and the formulas just established that

$$\frac{\partial \Gamma}{\partial j} = G_h / J_h = N \pm \frac{1}{J_h} \int_{z_1}^{z_c} \frac{dz}{\sqrt{n_o^2 - k^2 + 2h}} = w \quad (16.42)$$

$$\frac{\partial \Gamma}{\partial k} = k \frac{\partial G}{\partial h} + \frac{\partial G}{\partial k} = \sigma - \sigma^0$$

The last member in the first equation is our definition of  $w$ : it is the "angle variable" corresponding to the "action variable"  $j = J$ .

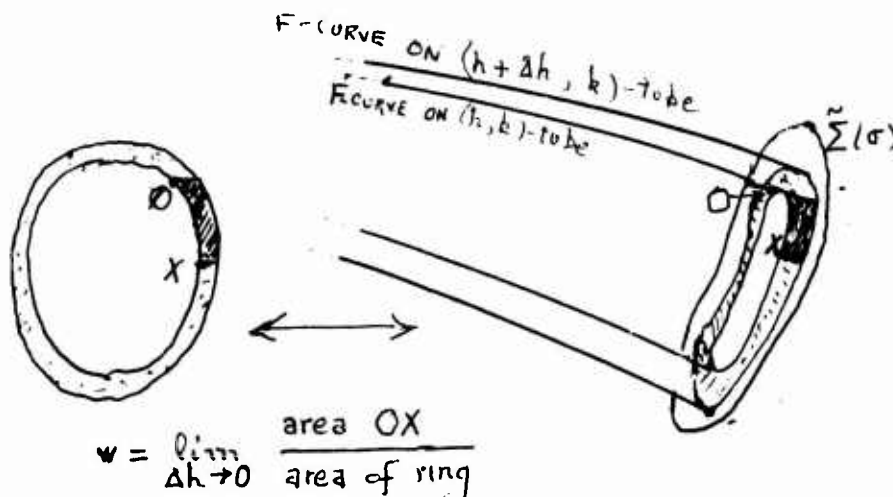


FIGURE 16.V

This variable  $w$ , which increases by an integer every time the path of integration  $\Lambda$  from  $B$  to  $X$  makes an additional clockwise loop around the tube, has a simple intrinsic definition, shown in Figure 16.V. Let the surface of section  $\Sigma(\sigma)$  be drawn through  $X$  and consider the ring between its two intersections with the  $(h, k)$  tube and the  $(h + \Delta h, k)$  tube. Through  $X$ , and also through the point  $O$  when  $\Sigma(\sigma)$  is cut by the  $F$ -curve through  $B$ , draw horizontal lines (i.e, lying in planes of constant  $z$ ). These will cut a piece  $OX$  from the ring. Divide the contour integral of  $\omega^*$  around this piece, by that around the whole boundary (two curves) of the ring--both in the same sense. The ratio will approach  $w$  as its limit as  $\Delta h \rightarrow 0$ . The proof consists in sliding the figure along  $F$ -curves into the initial plane  $r = r^0$ , using the invariance of the integrals. The multiplicity corresponds with the possibility of interpreting the "piece" as a multiple-covering figure, having the same extremities  $O$  and  $X$ .

We shall now apply the canonical transformation of (16.26) and (16.27), with the new momenta  $\bar{p}_1 = j$ ,  $\bar{p}_2 = k$ ,  $G$  being replaced by  $\Gamma = \Gamma(r, z, j, k)$ . In virtue of (16.42), our new conjugate position variables are  $\bar{q}_1 = w$  and  $\bar{q}_2 = \sigma - \sigma_0$ . By the general theory of canonical transformations, to find the new Hamiltonians for (16.1) and (16.24) replace the coordinates in  $H$  and  $F$  by their expressions in terms of the new variables: denoting the results by  $\bar{H}$  and  $\bar{F}$  we have, in virtue of (16.30).

$$\bar{H}(w, \sigma - \sigma_0, j, k) = h(j, k)$$

$$\bar{F}(w, \sigma - \sigma_0, j, k) = k$$

From these expressions the differential equations (16.1) become

$$(16.43) \left\{ \begin{array}{l} \frac{dw}{d\tau} = \frac{\partial \bar{H}}{\partial j} = \frac{\partial h}{\partial j} = 1/J_h \quad (\text{given in (16.36)}) \\ \frac{d\sigma}{d\tau} = \frac{\partial \bar{H}}{\partial k} = \frac{\partial h}{\partial k} = k \\ \frac{dj}{d\tau} = -\frac{\partial h}{\partial w} = 0, \quad \frac{dk}{d\tau} = -\frac{\partial h}{\partial w} = 0 \end{array} \right.$$

The last two are nothing new--they merely assert that  $j$  and  $k$  are first integrals. The first two show that  $w$  and  $\sigma$  increase at a constant rate, determined by the constant values of  $(j, k)$ .

The  $F$ -curves are the paths of translation parallel to the  $\sigma$ -axis, since their differential equations show that  $w, j, k$  are all constant along them, while  $\sigma$  increases. This is, of course, the same result as was obtained earlier, by the "deformation" based on the single equation (16.29). As before,  $\sigma$  plays the role of a variable of horizontal translation, carrying the class of  $H$ -curves for given  $(j, k)$  into itself, and each curve into itself whenever  $\sigma$  increases by the "ray period"  $\Pi(j, k)$ . And  $w$  acts as an angle variable, increasing constantly along each  $H$ -curve, and in terms of which the original variables are periodic functions of

period unity - as shown by (16.42) etc. Thus, while any point X of our phase space is determined by  $(w, \sigma, j, k)$ , X determines the latter only up to an additive integer for w.

After eliminating  $\tau$  from (16.43), its solution can, with the aid of (16.36), be written in terms of the function L defined as follows:

$$(16.44) \quad L = L(w, \sigma, j, k) = w - \frac{\sigma}{\Pi(j, k)} ;$$

(16.43) then shows that  $L(w, \sigma, j, k) = L(w^0, \sigma^0, j, k)$ . This means that L is another first integral of our H-equations: L, j, k form a set of three functionally independent first integrals -- the total possible number of independent integrals of the H-equations in the 4-dimensional phase space of the variables  $(r, z, p_r, p_z)$ .

We now return to the rays in the 3-dimensional phase space  $\mathfrak{P}_3$  for which  $h = 0$ . We lose one of our independent integrals since now  $j = J(0, k)$ . After introducing this expression in place of j in  $\Pi$  and L, we write  $\Pi = \Pi(k)$  and  $L = L(w, \sigma, k)$ . The two functions, k and L, furnish us with the maximum possible number of independent integrals, upon which any further first integrals must depend, by a relation such as (16.8).

We shall apply this fact to the acoustic power density quantity  $g = r p_r \tilde{f}(r, z, p_z)$  of (14.4), which was shown in Section 13 to be con-

stant along the rays -- a property obviously independent of the variables in terms of which it is expressed. We shall write  $g = g(w, \sigma, k)$ , and observe that it is periodic in  $w$ , with unit period. Being constant along the integral curves of (16.43) it is a first integral, and hence a function of  $(k, L)$ ; and we have

$$(16.45) \quad g(w, \sigma, k) = \Omega(k, L) = \Omega(k, w - \sigma/\Pi(k)),$$

the function of two variables  $\Omega$  being periodic in the second, with unit period. It is fully determined by the initial conditions, on  $\Sigma(r^0)$  where  $\sigma = \sigma^0 = r^0$ . There (16.45) shows that  $g(w, \sigma^0, k) = \Omega(k, w - \sigma^0/\Pi(k))$ , where  $(k, w^0)$  are the coordinates of a point on  $\Sigma(r^0)$ , with the "angular" property of  $w$ :  $(k, w + N)$  coinciding with  $(k, w)$ .

Let the initial values  $g(w, \sigma^0, k)$  be written as the function of two variables  $g_0(k, w)$ , so that, for all  $(k, w)$  on  $\Sigma(r^0)$  we have

$$\Omega(k, w - \sigma^0/\Pi(k)) = g_0(k, w)$$

Introducing the new variable  $x = w - \sigma^0/\Pi(k)$ , this becomes

$$\Omega(k, x) = g_0(k, x + \sigma^0/\Pi(k))$$

and this, being an identity in  $(k, x)$ , is valid when  $x = w - \sigma/\Pi(k)$ , whereupon the left member coincides with the right member of (16.45) —

which equation becomes

$$(16.46) \quad g(w, \sigma, k) = g_0(k, w - (\sigma - \sigma^0)/\pi k)).$$

We are assuming the initial values  $g_0(k, w)$  as given by the nature of the acoustic emission. Thus, the assumption made in Section 14 and applied ever since is that they have the constant value  $q$  on the injection region  $I$  and zero elsewhere on  $\Sigma(r^0)$ , this constant  $q$  being connected with the acoustic intensity on  $I$  by (14.8): this result was established at the emitter and requires no assumption of laminarity. This assumption can be written in terms of the characteristic function  $\chi_I(X)$  of the set  $I$ : equal to 1 or 0 according to whether the point  $X$  of  $\Sigma(r^0)$  is or is not on  $I$ . If  $X$  is given the coordinates  $(k, w)$ , we can write this function in the form  $\chi_I(k, w)$ , which must be periodic in  $w$ , of unit period. Then our earlier assumption takes the form

$$(16.47) \quad g_0(k, w) = q\chi_I(k, w).$$

This will be used in later calculations; but since our present analytical tools make it just as easy to deal directly with any more general  $g_0(k, w)$ , we shall carry out the derivation of the transmission loss formula first without using (16.47).

We turn now to the receiver: its range is  $r^1$  and depth  $z^1$ , lying in the depth interval  $(\zeta'', \zeta'')$ . We wish to calculate the acoustic intensity

at the point  $(r^1, z^1)$  of the ocean: the maximum power flux per unit ocean area, across an element of area  $\Delta S = \Delta C \cdot r^1 \Delta \phi$ , whose normal is at an angle  $\gamma'$  with the  $+z$  - axis. All directions of radiation (angles  $\gamma$ ) are included, i.e.,  $\gamma$  ranges over the interval  $\Gamma_1$  of all possible directions. Here  $\Delta C$  is an element of the curve  $C$  in the  $rz$  - plane, shown in Figures 13.III and IV. We recall the general formula (13.10) for the acoustic flux per unit azimuth,  $P_2$ , which was established under the sole assumption of azimuthal symmetry and not laminarity. Replacing  $C$  by  $\Delta C$  and  $\sigma$  by  $s$  (to avoid confusion with the present  $\sigma$ ), (13.10) becomes

$$P_2 = \int_{\Delta C} ds \int_{\Gamma_1} n(r, n \sin \gamma) \tilde{f} \cos \theta d\gamma$$

where  $\theta = \gamma - \gamma'$  is the angle between the fixed normal to  $\Delta C$  at the point  $(r^1, z^1)$  and the direction of radiation.

After the substitution  $r n \sin \gamma \tilde{f} = g$ , and the application of the law of the mean to the integration over  $\Delta C$ , we have

$$P_2 = \Delta C \int_{\Gamma_1} n g \cos \theta d\gamma,$$

where it is sufficiently accurate to assume that the space coordinates in the integrand are replaced by  $(r^1, z^1)$ ; we shall write, in particular,  $n = n_1 = n(r^1, z^1)$ . To obtain the acoustic intensity, note that  $P_2 \Delta \phi$  is the power flux through  $\Delta C$  in the azimuthal opening  $\Delta \phi$ , so we must divide this by the element of area  $\Delta C r^1 \Delta \phi$ ; then we must find the



direction  $\gamma'$  of maximum flux. Everything is as in the derivation (14.10) except for the fact that instead of  $qF(k)$  in the integrand we have the  $g$  given in (16.46) and in which the variables  $(w, \sigma, k)$  are regarded as functions of  $(r, z, p_z)$  (using corresponding superscripts), single-valued except for  $w$  which is determined only up to an additive integer, quite immaterial in determining the value of  $g$ . We have the following expression for the flux density in the direction  $\gamma'$ :

$$(16.48) \quad f_{l \cdot d}(\gamma') = \frac{1}{r^1} \int_{\Gamma_1} n_1 g_o(k, w - \frac{\sigma^1 - \sigma^0}{\Pi(k)}) \cos \theta \, d\gamma.$$

We are now in a position to lay down in precise mathematical terms our

Second defining assumption of secular non-laminarity:

In the above expression for the flux density, the factor  $g_o(\dots)$  in the integrand, which depends on the initial data and is periodic in  $\sigma^1 - \sigma^0$  with period  $\Pi(k)$ , may be replaced by its mean over a full period, namely by

$$(16.49) \quad \bar{g}_o(k) = \frac{1}{\Pi(k)} \int_0^{\Pi(k)} g_o(k, w - \frac{\sigma^1 - \sigma^0 + s}{\Pi(k)}) \, ds.$$

This is an assumption of ray-phase incoherence, and reflects, as did our various pictures of "ergodic mixing", the fact that the periods  $\Pi(k)$  vary enough with  $k$  to obliterate, after a few convergence zones, any precise phase relations in the initial data.

The fact that  $\bar{g}_0$  depends only on  $k$  follows from the property of periodic functions averaged over any full period, and which can be made evident by the change of variable of integration  $x = (\sigma^1 - \sigma^0 + s)/\pi(k)$ , which converts the integral in (16.49) into

$$\int_{x_1}^{x_1+1} g_0(k, x) dx;$$

$g_0(k, x)$  is periodic in  $x$  with period unity and  $x_1 = (\sigma^1 - \sigma^0)/\pi(k)$ : simply differentiate this integral with respect to  $x_1$ ; the result is  $g_0(k, x_1 + 1) - g_0(k, x_1) = 0$ .

The mean  $\bar{g}_0(k)$ , with assumption (16.47), becomes the function of  $qF(k)$  of Section 14 and based on (14.6). This fact could be shown as previously by the ideas of the mixing of dyed regions—whose initial intensity of dye is specified by the second member of (16.47). But we do not need any such further appeal to intuition: our second assumption stated above in the replacement of  $g_0$  in (16.48) by  $\bar{g}_0$  of (16.49) is all that is required.

With this assumption (16.48) becomes

$$(16.50) \quad \text{fl} \cdot d \cdot (\gamma') = \frac{1}{r^1} \int_{\Gamma_1} n_1 \bar{g}_0(k) \cos \theta \, d\gamma$$

Now this coincides with (14.10)' when  $\bar{g}_0(k)$  is replaced by  $qF(k)$ , and vice versa. As in Section 14, the domain  $\Gamma_1$  over which the angle  $\gamma$  is integrated, and which we have assumed to contain all possible directions of rays within a duct in an ocean with at most a slowly varying bottom depth  $b$  (i.e., imperceptibly changing over a convergence zone distance), is easily seen to be symmetric about its center ( $\gamma = \pi/2$ ); therefore, the vertical component in (16.50) vanishes; the radial component alone giving the acoustic intensity: just as (14.10)' led to (14.11), (14.12), and to the transmission loss expression (14.13), so we can assume that these equations apply to the present case — after the symbol  $qF(k)$  is replaced by  $\bar{g}_0(k)$ .

It remains to find a convenient expression of the  $\bar{g}_0(k)$  defined by (16.49) in terms of the initial values  $g_0(k, w)$  on  $\Sigma(r^0)$ . This is easily done by applying the basic facts concerning the integral invariants in Hamiltonian systems, set forth in Appendix E and using as coordinates in  $\mathcal{H}_3$  our variables  $w, \sigma, k$ . We have seen that both  $g(w, \sigma, k)$  and the  $L(w, \sigma, k)$  defined by (16.44) are first integrals of the ray equations; therefore by Cartan's theorem the product  $g\delta L$ , being expressible exclusively in terms of such first integrals and their differentials, is the integrand of an absolute sliding integral invariant. We are interested in it on the manifolds  $k = \text{constant}$  (cf. the "Snell tubes" of Figure 16.I(a) with  $h = 0$ ). On each of these,  $\delta k = 0$ , and we have by (16.46)

$$(16.51) \quad g\delta L = g_0(k, w - (\sigma - \sigma^0)/\Pi(k)) \cdot (\delta w - \delta\sigma/\Pi(k)).$$

Let us integrate this expression with respect to  $\sigma$  on the line segment from  $\sigma = \sigma^0$  to  $\sigma = \sigma^0 + \Pi(k)$ , ( $w$  and  $k$  constant). This is a piece of an element of the cylindrical Snell tube—cylindrical, in the space of our new variables. Since in the integrand  $\delta w = 0$ , the resulting integral becomes precisely  $-\bar{g}_0(k)$  defined in (16.49) (with the variable of integration  $\sigma = \sigma^1 + s$ , an immaterial change, when the integration is over a whole period).

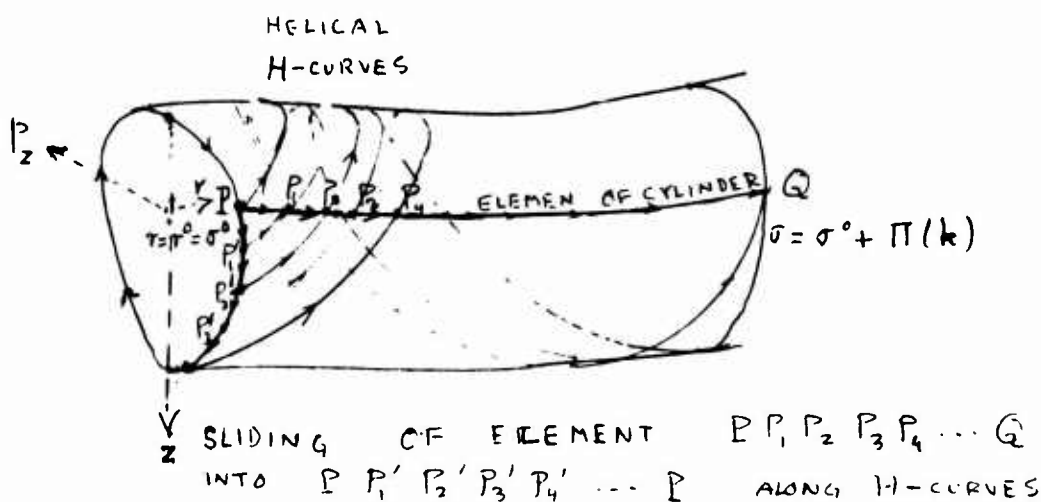


FIGURE 16.VI

Now slide the above path of integration along the H-curves into the plane  $\Sigma(r^0)$ ; see Figure 16.VI. Since these stay on the cylinder, the new path lies along its intersection with the initial plane  $\Sigma(r^0)$ ; and since its original  $\sigma$ -length was the period  $\Pi(k)$ , it becomes a closed curve in  $\Sigma(r^0)$ , traced exactly once in the positive or counterclockwise sense, as is seen in the figure, inasmuch as the rays themselves wind

clockwise around the cylinder.

By the invariance of the integral, its value can be calculated in the  $\Sigma(r^0)$  plane, where  $\delta\sigma=0$  and  $\sigma = \sigma^0$ , which leads by (16.51), to the contour integral

$$-\bar{g}_0(k) = \int_0^1 g_0(k, w) \delta w$$

On returning to the old variables  $(z, p_z)$  in the plane  $\Sigma(r^0)$ , with  $w$  replaced by its expression in (16.42), we obtain the result which, when—as can always be safely assumed— $g_0$  is symmetric in  $p_z$ , reduces to

$$(16.52) \quad \bar{g}_0(k) = \frac{2}{\pi(k)} \int_{r_1(k)}^{r_2(k)} g_0[z, p_z] \frac{dz}{\sqrt{n_0^2 - k^2}},$$

where  $g_0[z, p_z] = g_0(k, w)$ , representing the initial values in terms of the old coordinates. And when in particular this is given by (16.47), with  $\chi_I(k, w) = \chi_I[z, p_z]$ , we see that the integrand in (16.52) is zero except in the depth interval  $(z', z'')$ , the width of  $I$ . If, as in Section 14,  $z^0$  is an appropriately chosen intermediate depth, the law of the mean converts (16.52) into the form

$$\bar{g}_0(k) = q(z'' - z') \frac{2k}{\sqrt{n^2(z_0) - k^2}} \frac{1}{\Pi(k)} = qF(k),$$

the last, in virtue of (14.17), since  $R(k) = \Pi(k)$ .

From this point on, we merely replace  $\bar{g}_0(k)$  by the above value in (16.50) to obtain, passing through (14.12) and (14.13), the transmission loss formulas (14.19) and (14.20), with  $R(k)$  replaced by  $\Pi(k)$ .

#### Concluding Remarks

A. The final (mathematical) version of the method used is founded on six bases: (1) the identification of geometrically simple first integrals of the Hamiltonian equations governing the rays in the unperturbed (laminar) case; (2) the extension of their existence and simple geometrical nature in the perturbed case, in the ranges corresponding to the physical problem of interest, by the theorems of analysis of the G.D. Birkhoff type; (3) the introduction of as many Hamiltonian systems (all in the same phase space of canonical coordinates) as there are independent geometrically simple first integrals of the given (perturbed) system; (4) the construction of a simultaneous solution of the corresponding

set of Hamilton-Jacobi equations by a line integral, the validity and properties of which—over the long ranges of importance—is based on E. Cartan's theory of sliding integral invariance; (5) the use of this solution to change canonically to angle and action variables; (6) the application of the simple (linear or periodic) form of the ray equations in these coordinates to acceptable approximations of the transmission loss, based on mathematically precise and physically meaningful assumptions.

B. The second defining property of secular non-laminarity could have been based on the possibility of a Fourier series expansion of our periodic functions involving the variable  $w$ : the approximation  $\bar{g}_0$  of (16.49) would then appear as the constant term in the expansion of the general density  $g$ . The reason for dropping the non-constant terms is the "smearing" of periods  $\Pi(k)$  in the terms in  $\exp 2\pi i[w - \sigma/\Pi(k)]$  as  $k$  varies—the familiar picture of "mixing" applied so often.

C. The method has three obvious extensions: first, the one to the non-azimuthally symmetric case  $c = c(x, y, z)$  noted before. Second, to the "random transformation" conception of the flow on the surface of section with the application of information theory—doing for the secularly non-laminary cases what was done in Section 15 for the laminar, and thus achieving a great simplification at long ranges. Third, to the case of two ducts in the laminar and the perturbed situation: there will be a different sort of multiple-valuedness, setting in at certain critical values of the small constant  $k$  (cf. the "figure eight" Snell curves depicted in Figure 14.IV).

D. The case of duct splitting, having no laminar analogue, cannot have geometrically simple first integrals established by the perturbation of a laminar situation. This does not exclude the possibility of directly studying the topology of the loci of  $P_r = k$  for the  $P_r$  defined as in (16.11), but with the condition  $h = 0$  set aside. To obtain an effective hold on the problem we can assume, for the reasons given early in Appendix C, that the coefficients (essentially  $c$ ) are analytic in the coordinates for a region  $R$  of the space of complex values of the latter; and that the physical space  $W$  of the acoustic transmission and which corresponds to their real values, is embedded in the interior of  $R$ . Then the integral  $P_r$  is itself analytic within the latter region, and the study of duct doubling is that of the behavior of the solutions of  $P_r = k$  with changing  $r$ : the appearance or disappearance of solutions is interpreted in terms of pairs of conjugate imaginary solutions (that were there all along!) becoming real, and vice versa. The introduction of angle and action variables by the line integral solution of the simultaneous Hamilton-Jacobi equations, and their Fourier expansion has to be replaced by a more sophisticated uniformization of the loci  $H=h$ ,  $P_r = F = k$ , by line integrals in the complex domain  $R$  -- all which ideas have their roots in the uniformization of algebraic loci by elliptic and hyperelliptic integrals, of Hodge's theory of harmonic integrals on a real manifold, etc.



E. The method used in this Section is a new combination of a number of quite diverse classical fragments. Apart from its applications to acoustic propagation in heterogeneous media, it has many other potential applications — to classical dynamics, to cite the most obvious one.

## REFERENCES

1. General works on propagation of sound and similar disturbances:

P. M. Morse & K. U. Ingard, "Theoretical Acoustics", (McGraw-Hill, New York, 1968).

L. M. Brekhovskikh, "Waves in Layered Media", (Academic Press, New York, 1960).

C. B. Officer, "Introduction to the Theory of Sound Transmission", (McGraw-Hill, New York, 1958).

I. Tolstoy & C. S. Clay, "Ocean Acoustics", (McGraw-Hill, New York, 1966).

2. Works emphasizing the random factors in the above types of propagation:

L. A. Chernov, "Wave Propagation in a Random Medium", (McGraw-Hill, New York, 1960).

V. J. Tatarski, "Wave Propagation in a Turbulent Medium", (McGraw-Hill, New York, 1961).

M. J. Beran, J. J. McCoy, B. B. Adams, "Effects of a Fluctuating Temperature Field on the Spatial Coherence of Acoustic Signals", NRL Report 7809, May 22, 1975.

As but a few examples of the many publications in the learned journals dealing with this subject, may be cited:

H. W. Marsh, "Sound Reflection and Scattering from the Sea Surface", The Journal of the Acoustical Society of America, Vol. 35, No. 2, February 1963. Also cf. the scores of papers in the above Journal on this subject that appeared since that date; in particular, M. A. Biot, "Generalized Boundary Conditions for Multiple Scatter in Acoustic Reflection", Vol. 44, (1968), D. S. Bugnolo, "Stochastic Born Series for Sound Propagation in Random Media", Vol. 44, (1968), and I. Dyer, "Statistics of Sound Propagation in the Ocean", Vol. 48, (1970).

3. General works dealing with the mathematical background of wave propagation in acoustics and electromagnetism:

J. Hadamard, "Lecons sur la Propagation des Ondes", (Herman, Paris, 1903).

J. Hadamard, "Lectures on Cauchy's Problem in Linear Partial Differential Equations", (Yale University Press, New Haven, 1923).

R. Courant & D. Hilbert, "Methods of Mathematical Physics", Vols. I and II. (Interscience, New York, 1966).

J. A. Stratton, "Electromagnetic Theory", (McGraw-Hill, New York, 1941 and later editions).

4. Works on calculus and differential geometry:

W. F. Osgood, "Advanced Calculus", (McMillan, New York, 1925).

E. Goursat & E. R. Hedrick, "Mathematical Analysis", (Originally, Ginn & Co., then Dover Publications, Inc., New York, 1924).

L. P. Eisenhardt, "Treatise on the Geometry of Curves and Surfaces", Ginn & Co. (Boston, 1909).

W. Blaschke, "Differentialgeometrie", Vol. I, (Julius Springer, Berlin, 1921).

5. Integral invariants in Hamiltonian systems:

E. Whittaker, "Analytical Dynamics", Cambridge at the University Press (Second Edition, 1917).

H. Poincaré, "Méthodes Nouvelle de la Mécanique Céleste", Vol. III, (Gauthier-Villard, Paris, 1899).

E. Cartan, "Lecons sur les Invariants Intégraux", (Hermann, Paris, 1922 and 1958).

6. Modern dynamics, ergodic theory, and information theory:

J. W. Gibbs, "Collected Works", Vol. II, (Longmans Green and Co., New York, 1928).

G. D. Birkhoff, "The Restricted Problem of Three Bodies", Rendiconti del Circolo Matematico di Palermo (23 Aug. 1914).

G. D. Birkhoff, "Dynamical Systems", Amer. Math. Soc. Colloquium Publications, Vol. IX (1927).

G. D. Birkhoff & B. O. Koopman, "Recent Contributions to Ergodic Theory", Proc. Nat. Acad. of Sci., Vol. 18 (March 1932), pp. 279-282.

B. O. Koopman & J. V. Neumann, "Dynamical Systems of Continuous Spectra", Proc. Nat. Acad. of Sci., Vol. 18 (March 1932), pp. 255-263.

P. R. Halmos, "Lectures on Ergodic Theory", Chelsea Publishing Co., New York, N.Y. (1956).

S. Kullback, "Information Theory and Statistics", (Wiley, New York, 1959).

## APPENDIX A

### THE MATHEMATICS OF RAYS AND WAVE FRONTS: CERTAIN COMMON ERRORS

The starting point in the ray treatment of propagation--for both underwater sound and geometrical optics--is Fermat's Principle.<sup>\*</sup> This states that the path of propagation, the ray L, has the property that if any two points A and B are picked on L and if they are not "too far apart", the arc AB of L is the path of least time of travel of the disturbance from A to B, in comparison with the time taken along any other curve joining A to B. Since the travel time  $T_{AB}$  is given by the following line integral along the curve from A to B

$$(A.1) \quad T_{AB} = \int_A^B ds/c = \int_A^B n ds,$$

(where  $ds$  is the differential of arc length) the basic property of the ray is that it minimizes the integral in (A.1).

The application of the elements of the calculus of variations leads to the Euler differential equations.\*\* Their form depends on the coordinate system. In rectangular coordinates  $(x,y,z)$ , the element of arc is given by the formula  $ds^2 = dx^2 + dy^2 + dz^2$ , and the Euler equations are as follows, the travel time  $t$  being the independent variable:

$$(A.2) \quad \frac{d}{dt}(n^2 \frac{dx}{dt}) = \frac{1}{n} \frac{\partial n}{\partial x}, \quad \frac{d}{dt}(n^2 \frac{dy}{dt}) = \frac{1}{n} \frac{\partial n}{\partial y}, \quad \frac{d}{dt}(n^2 \frac{dz}{dt}) = \frac{1}{n} \frac{\partial n}{\partial z}$$

The elementary theorems applying to such equations tell us that they have solutions expressing the coordinates of the point

<sup>\*</sup>See Officer,<sup>1</sup> p. 42; Tolstoy & Clay,<sup>1</sup> p. 53.

<sup>\*\*</sup>cf. Osgood,<sup>4</sup> Chapter XVII.

$(x,y,z)$  (moving on the ray) as functions of the independent variable  $t$  (the travel time along the ray in question). They also tell us that through any given point  $(x_0,y_0,z_0)$  and tangent to any initial direction through it, specified, e.g., by angular parameters  $(\theta,\phi)$ , there passes one and only one ray. In other words, there exist three functions giving the solution of (A.2) in the form

$$(A.3) \quad \begin{aligned} x &= f(t, x_0, y_0, z_0, \theta, \phi) \\ y &= g(t, x_0, y_0, z_0, \theta, \phi) \\ z &= h(t, x_0, y_0, z_0, \theta, \phi) \end{aligned}$$

As  $t$  varies (the other five variables being held fixed) the point with the above coordinates moves along the ray determined by the given values of the five parameters, starting at  $(x_0,y_0,z_0)$  in the direction specified by  $(\theta,\phi)$  when  $t = 0$ .

It is emphasized that the existence and various other properties of these solutions upon which we shall base our reasoning and computational methods does not imply that the functions in (A.3) can necessarily be expressed by simple formulas in terms of standard functions: the restrictive conception of "analytic treatments" to the cases where this is possible has been discarded by mathematics centuries ago! The solutions (A.3) can always be computed numerically, to the order of accuracy with which  $n = 1/c$  is known; with modern computers this process can be done expeditiously. This fact is the basis of the ray tracing technique referred to in Section 1.

The object here is to scrutinize the scientific validity of the process and of the applications of it that are usually made--in the light of the mathematical facts of the whole situation. Explicit calculations by practicable methods is a different subject.

When two points A and B are given and are "not too far apart" (in a certain mathematically precise sense), one and only one ray passes between them. Accordingly, the coordinates  $(x, y, z)$  of the point tracing this ray could have been given by equations like (A.3),  $(x_0, y_0, z_0)$  being the coordinates of A, but with  $(\theta, \phi)$  replaced by  $(x_1, y_1, z_1)$  the coordinates of B.

In spite of the six parameters in the latter case and the five in the former, different sets of values of the parameters do not necessarily give different rays. Actually the totality of rays is a four parameter family. This is evident in the case  $n = \text{constant}$ , when the rays are straight lines: If we take a pair of parallel reference planes, the set of lines cutting both planes is in one-to-one correspondence with pairs of points, one in each plane, and hence with their two pairs of coordinates. The exceptional lines, which are parallel to the reference planes, form a set depending on only three parameters, as a simple construction shows.

The second basic fact emerging from the standard methods of calculus of variations is the characterization of the wave fronts  $W_0$ ,  $W$ , etc. of Figures 1.1 and 1.4 as surfaces of constant phase.<sup>\*</sup> This means that the travel time  $T_{Op_0}$  along the ray from O to  $P_0$  is the same for

---

<sup>\*</sup>cf. Officer,<sup>1</sup> p. 45.

all other rays from  $O$  to points  $P'_0$  on  $W_0$ ; and similarly for the travel times from  $O$  to the various points of  $W$ . Therefore, by subtraction, the travel time along the ray from  $P_0$  to  $P$  is the same as along any other ray (of our bundle through  $O$ ), from  $P'_0$  to  $P'$ , joining the wave fronts  $W_0$  and  $W$ . Finally, each of these wave fronts, like  $W$ , cuts all the rays through  $P$  at right angles. Putting these properties together, we see that whether we define the wave fronts as surfaces of constant phase from  $O$  or as orthogonal surfaces to the rays through  $O$ , we get the same surfaces  $W$ . (Law of Malus and Huygen's construction).

We return now to the basic problem of Section 1: to find the spreading factor, the ratio of areas  $S/S_0$  as  $S \rightarrow P$  and  $S_0 \rightarrow P_0$ . We give the general principles here, deferring more detailed calculations to the next section.

In equations (A.3), think of  $(x_0, y_0, z_0)$  as the coordinates of the point of emission  $O$ . This could be taken as the origin of our coordinate system, so that these three numbers would become  $(0, 0, 0)$ ; but this is not essential. What is essential is to think of  $(x_0, y_0, z_0)$  as held fixed. Then (A.3) express the 2-parameter bundle of rays issuing through  $O$ , each characterized by its two direction parameters,  $\theta, \phi$ . If one wishes an explicite geometrical interpretation, one can choose these as the co-latitude and longitude of the intersection with a sphere (of unit radius and centered at  $O$ ) of the tangent to the ray at  $O$ . In this construction it is usual to take the north pole on the



+z-axis and the initial meridian ( $\phi=0$ ) in the xz-plane. In conclusion, if we drop the fixed  $(x_0, y_0, z_0)$  from (A.3) we obtain the equations of our 2-parameter bundle of rays as

$$(A.4) \quad x = f(t, \theta, \phi), \quad y = g(t, \theta, \phi), \quad z = h(t, \theta, \phi)$$

We have seen that for all  $(x, y, z)$  on the wave front  $W$ ,  $t$  has a constant value. Denote its value for  $W$  by  $t_1$ , and for  $W_0$  by  $t_0$ . Then the equations

$$(A.5) \quad x = f(t_i, \theta, \phi), \quad y = g(t_i, \theta, \phi), \quad z = h(t_i, \theta, \phi)$$

are the parametric equations of the surface  $W$  when  $i = 1$ , and of  $W_0$  when  $i=0$ .<sup>\*</sup> The "parameters" are the angular variables  $(\theta, \phi)$ . In the special case of Figure 1.I, they can be taken as the co-latitude and longitude; but in the general case of Figure 1.II, the wave fronts are not spheres and the angles have to be regarded merely as general curvilinear (Gaussian) coordinates on the surface in question. In all cases, the element of area ( $\delta S$  or  $\delta S_0$ , say) on the wave fronts  $W$  and  $W_0$  can be expressed explicitly in terms of (A.5).

On the surface  $W$  (given by (A.5) with  $i = 1$ ), when  $\theta$  is kept fixed and  $\phi$  allowed to vary, the point  $(x, y, z)$  traces a curve (called a  $\phi$ -curve); it will trace another nearby  $\phi$ -curve when the fixed value  $\theta$  is replaced by a neighboring fixed value  $\theta + \delta\theta$ . Similarly when  $\phi$  is held fixed at one value, and then at a neighboring value  $\phi + \delta\phi$ , and  $\theta$

---

<sup>\*</sup> Cf. Officer<sup>1</sup>

varies, we obtain two nearby  $\theta$ -curves. These are shown in Figure A1 which pictures the curvilinear quadrilateral  $\Delta S$  bounded by these four curves. It also shows a "tangent" quadrilateral  $\delta S$  which is a parallelogram and is bounded by straight line segments -- namely the vectors having the following components (subscripts denoting partial derivatives):

$$(x_\phi \delta\phi, y_\phi \delta\phi, z_\phi \delta\phi) \quad (x_\theta \delta\theta, y_\theta \delta\theta, z_\theta \delta\theta)$$

tangent to the  $\theta$ - and  $\phi$ -curves. Areas  $\Delta S$  and  $\delta S$  differ by quantities of higher order (so that  $\Delta S/\delta S \rightarrow 1$ ). The area of the latter parallelogram is the product of its two adjacent sides with the sine of the angle between them; therefore it is the absolute value of the vector product of its two vector sides. Therefore

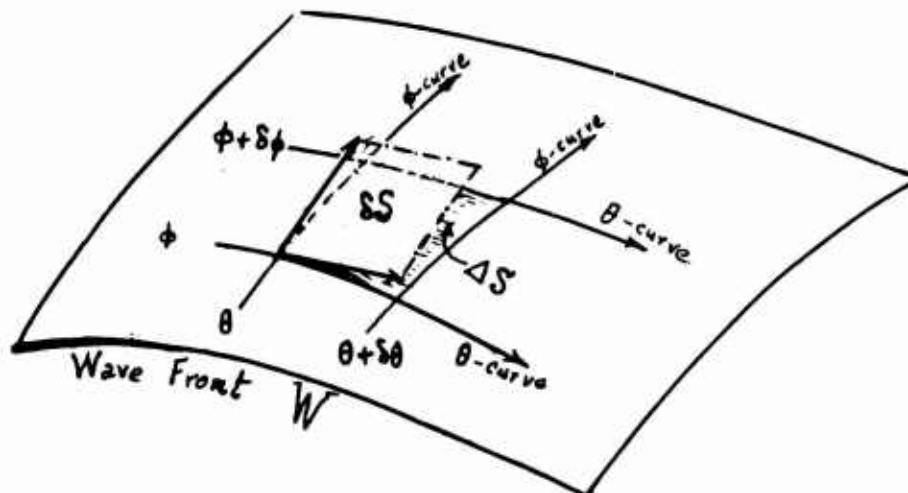


Figure A1      Element of Area on W and Tangent Parallelogram

$$(A.6) \quad \delta S = |\vec{r}_\theta \times \vec{r}_\phi| \delta\theta \delta\phi \\ = [(y_\theta z_\phi - z_\theta y_\phi)^2 + (z_\theta x_\phi - x_\theta z_\phi)^2 + (x_\theta y_\phi - y_\theta x_\phi)^2]^{1/2} \delta\theta \delta\phi;$$

in the first line, vector notation is used,  $\vec{r}$  being the vector of components (x,y,z), and subscripts again denoting partial differentiation.

Since the required spreading factor is the limit of the ratio  $\Delta S/\Delta S_0$  it is equal to that of  $\delta S/\delta S_0$ . Hence, by (A.6) applied first with  $t = t_1$  and then with  $t = t_0$ , we obtain

$$(A.7) \quad \text{Spreading factor } \delta S/\delta S_0 = \frac{|\vec{r}_\theta \times \vec{r}_\phi| (t=t_1)}{|\vec{r}_\theta \times \vec{r}_\phi| (t=t_0)}$$

This vector form has not only the advantage of compactness, but of being independent of the coordinate system, a fact which we will use in the next appendix.

Thus we have reduced the problem of finding the transmission loss to that of evaluating the quantity on the right in (A.7), which evaluation will be provided as soon as the solution (A.4) of (A.2) is obtained. The practical question of what is the most economical computation to find it numerically is one issue facing us; but it is only one: the over-riding issue is whether the assumptions which have underlain the above discussion are correct in the important cases. The assumptions have been two in number: first, that n (or c) is fixed in t and known throughout the region of propagation; second, the absence of multiple-path transmission, intersecting wave fronts, and caustics, which are examined in Section 2.

We conclude this introductory discussion of Fermat's principle, rays, wave fronts, and spreading factors by examining certain deductions from this principle known more or less loosely as Snell's law, and certain pitfalls its misinterpretation have sometimes caused. And we shall also note a rather frequent fallacy in constructions of shadow zones by "critical rays".

Snell's law and its misinterpretation. It is convenient to re-write (A.2) in terms of the arc length  $s$ , measured along the ray in the direction of propagation, as the independent variable instead of  $t$ , using  $ds = c dt$ ,  $dt = nds$ , and the direction cosines of this ray direction, given by  $dx/ds = \cos \alpha$  etc. The equations become

$$(A.8) \quad \frac{d}{ds} (n \cos \alpha) = \frac{\partial n}{\partial x}, \quad \frac{d}{ds} (n \cos \beta) = \frac{\partial n}{\partial y}, \quad \frac{d}{ds} (n \cos \gamma) = \frac{\partial n}{\partial z}.$$

Suppose that  $n$  is independent of  $x$ ; the first of the above equations shows that  $n \cos \alpha$ --explicitly,  $n(y,z) \cos \alpha(y,z)$ --is constant along each ray: for reasons that will appear, this fact may be called Snell's law throughout the medium--with respect to directions parallel to the  $x$ -axis. Conversely, when  $n \cos \alpha$  is constant along each ray,  $\partial n / \partial x = 0$ , so that  $n$  is independent of  $x$ . Similarly for the other variables.

A usual form of Snell's law throughout the medium occurs when it is "layered", i.e., when  $n$  is independent of two coordinates, e.g.,  $x$  and  $y$ :  $n = n(z)$ . From the constancy of  $n \cos \alpha$  and  $n \cos \beta$ , and the relations between direction cosines, we have

$$n^2 \sin^2 \gamma = n^2 (1 - \cos^2 \gamma) = n^2 \cos^2 \alpha + n^2 \cos^2 \beta$$

which is constant, and similarly for  $n \sin \gamma$ . Since  $\gamma$  is the angle between the ray and the normal to the layers (i.e., to the gradient of  $n$ ), the fact that at two points  $P_1$  and  $P_2$  on the same ray,  $n_1 \sin \gamma_1 = n_2 \sin \gamma_2$ , expresses Snell's law in its familiar form in terms of the angles of incidence at  $P_1$  and of refraction at  $P_2$  in crossing the layer of the medium between these points. Conversely, suppose that  $n \sin \gamma$  is constant along each ray. Differentiating the above equation and applying (A.8), etc., we obtain

$$\cos \alpha \frac{\partial n}{\partial x} + \cos \beta \frac{\partial n}{\partial y} = 0;$$

and this being true for each ray direction (i.e., regarding  $\cos \alpha$  and  $\cos \beta$  as independent variables) we derive the identities  $\partial n / \partial x = 0$  and  $\partial n / \partial y = 0$ , so that  $n$  depends only on  $z$ . Thus, Snell's law throughout the medium in the above sense is valid if and only if the medium is layered.

In the discussion so far, we have assumed that the properties of the medium vary "smoothly" from point to point: that  $n$  together with its derivatives through the second order are continuous: this (or a little less) will insure the derivability of (A.2) or (A.8) from Fermat's principle (A.1). We now examine a different case, that of two media separated by a smooth interface  $\Sigma$ , each individually smooth in the above sense, up to and inclusive of  $\Sigma$ , but experiencing a "finite jump" across  $\Sigma$ : at each point  $X$  on  $\Sigma$ , the limit  $n_1$  of  $n$  as  $X$  is approached from one side is not equal to the limit  $n_2$  from the other side, the jump

being  $n_2 - n_1$ . It is then possible to prove directly from Fermat's principle (A.1) that for any ray through P (i.e., time-minimizing path AXB),  $n_1 \sin \theta_1 = n_2 \sin \theta_2$ , where  $\theta_1$  and  $\theta_2$  are the angles that the two pieces of the ray make with the normal to  $\Sigma$  at X (the angles of incidence and of refraction). This deduction from Fermat's principle does not require  $n$  to be independent of any variable of position: it can be the most general function, smooth except at the interface  $\Sigma$ . This may be called Snell's law at an interface.

The proof makes use of classical methods of the calculus of variations, sufficiently illustrated by the case of propagation in the  $xy$ -plane and with the interface  $\Sigma$  (now 1-dimensional) taken as the  $x$ -axis. Let A be a point in the first medium ( $y < 0$ ) and B in the second ( $y > 0$ ), and let the coordinates of X be  $(\xi, 0)$ . To minimize the time  $T_{AB}$  along the path AXB, we must evidently take for the two pieces AX and XB the rays in each medium connecting A with X and X with B (uniquely determined when these three points are, as we are assuming, close enough together). Each piece of ray will satisfy the Euler equations--and we shall now take  $x$  as the independent variable.

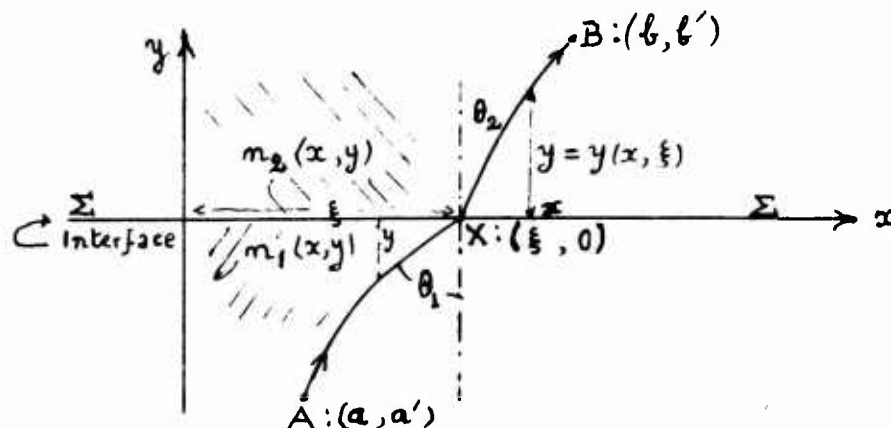


Figure A-II. Interface Relations

Evidently  $y = y(x, \xi)$ : as  $x$  varies and  $\xi$  stays fixed,  $(x, y)$  moves along the ray; as  $\xi$  varies while  $x$  is fixed,  $(x, y)$  moves vertically from ray to ray. It is this function  $y(x, \xi)$  which must satisfy Euler's equation (single, in this case) where  $x$  is the independent variable and  $\xi$  a fixed parameter. We use the notation  $y' = \partial y / \partial x$ ,  $y^* = \partial y / \partial \xi$ .

The time of propagation along AXB, when AX and XB are rays, will depend on the position of X, i.e., on the value  $\xi$ :

$$T_{AXB} = \int_a^\xi n ds + \int_\xi^b n ds = T(\xi).$$

To find its minimum we set the derivative  $T'(\xi) = 0$ . We may apply Leibnitz' rule to each term (differentiation under the integral sign), adding the terms corresponding to the variable limits. Thus from the first term,

$$\int_a^\xi n ds = \int_a^\xi n(x, y(x, \xi)) [1 + y'(x, \xi)^2]^{1/2} dx$$

we obtain

$$\begin{aligned} \frac{d}{d\xi} \int_a^\xi n ds &= \int_a^\xi \frac{\partial}{\partial \xi} \{ n [1 + (y')^2]^{1/2} \} dx \\ &\quad + n_1(\xi, 0) [1 + (y'_1)^2]^{1/2} \end{aligned}$$

Here the subscript 1 means the limit of the function as the point X is approached from below, while the accent on  $y$  denotes the derivative with respect to  $x$ ,  $\xi$  being held fast.

The  $\xi$ -derivative under the integral is worked out, the usual integration by parts applied to the term containing  $\partial y' / \partial \xi = \partial(\partial y / \partial \xi) / \partial x$ ; and the coefficients of  $\partial y / \partial \xi$  gathered together: their sum is zero, by

Euler's equation for AX. There remains the integrated part, which, together with the term contributed by the variable upper limit, yields

$$n_1(\xi, 0) \left\{ \frac{y_1'(\xi, \xi) y_1^*(\xi, \xi)}{\sqrt{1 + y_1'(\xi, \xi)^2}} + \frac{1}{\sqrt{1 + y_1'(\xi, \xi)^2}} \right\}.$$

Since along the x-axis we have  $y(x, \xi) = y(\xi, \xi) = 0$ , differentiation by  $\xi$  shows that  $y'(\xi, \xi) + y^*(\xi, \xi) = 0$ . Replacing  $y^*$  by  $-y'$  in the above expression reduces it to  $n_1(\xi, 0) [1 + y_1'(\xi, \xi)^2]^{-1/2}$  which is  $n_1(\xi, 0) \cos \alpha_1$ ; or, using the angle of incidence  $\theta_1$ , to  $n_1 \sin \theta_1$  at X.

A precisely similar treatment applies to the second term,  $T_{XB}$ , the result being  $-n_2 \sin \theta_2$ . Since the derivative of  $T_{AXB}$  must be zero for the minimum, we have  $n_1 \sin \theta_1 = n_2 \sin \theta_2$ , Snell's law at an interface: valid on  $\Sigma$  only.

The confusion observed too often in memoranda and even in some published works is between Snell's law at an interface and Snell's law throughout the medium: the first, holding only at the lower dimensional interface figure  $\Sigma$ , but applying to any smooth interface between smooth but general media; the second applying throughout the medium but valid only in the special cases of independence of the index on spacial displacements along one or more directions. Knowing that the interface law is perfectly general, the pitfall lies in assuming its validity throughout the medium. In view of what has been established above, reasoning based on such an assumption is incorrect: there is no possibility of extending the interface property to the whole general medium, e.g., by approximating the latter by thin homogenous layers,



etc.--in spite of the analogy with the continuous-versus lumped constants methods that succeed in many other physical situations.

The "critical ray" fallacy. Another instance of wrong reasoning due to "lumping" continuous quantities (in this case, velocity gradients) in ray propagation occurs when the acoustic profile of the medium--which we now suppose to be layered, with  $c$  (and  $n$ ) dependent on depth  $z$  only--is approximated by two or more straight lines. Ever since its earliest days, the calculus of variations has shown that in any interval  $(z_1 < z < z_2)$ , where  $c$  is linear the rays are arcs of circles. Furthermore, Fermat's principle shows that at junction points of two linear representations of  $c$  (e.g.,  $z_2$  if between  $z_2$  and  $z_3$  another line is used, the two forming a connected broken line between  $z_1$  and  $z_3$ ) the corresponding circular rays will be mutually tangent. This leads to a greatly simplified construction of rays: one merely draws a set of consecutive tangent circular arcs, each circle having its radius determined by the slope of its initial tangent and that of the linear representation of  $c$  (cf. (D.7) in Appendix D). Therefore it leads to a graphical evaluation of the spreading factor and thus of the transmission loss, by the reasoning of Section 1. It gives, for example, the construction of shadow zones, the "critical ray", etc., shown in Figure A-III, when at the junction point  $z_2$   $c$  has a maximum (as under a layer). Except at ranges such that other rays enter it, e.g., by reflection, no acoustic power from A enters the shadow zone.

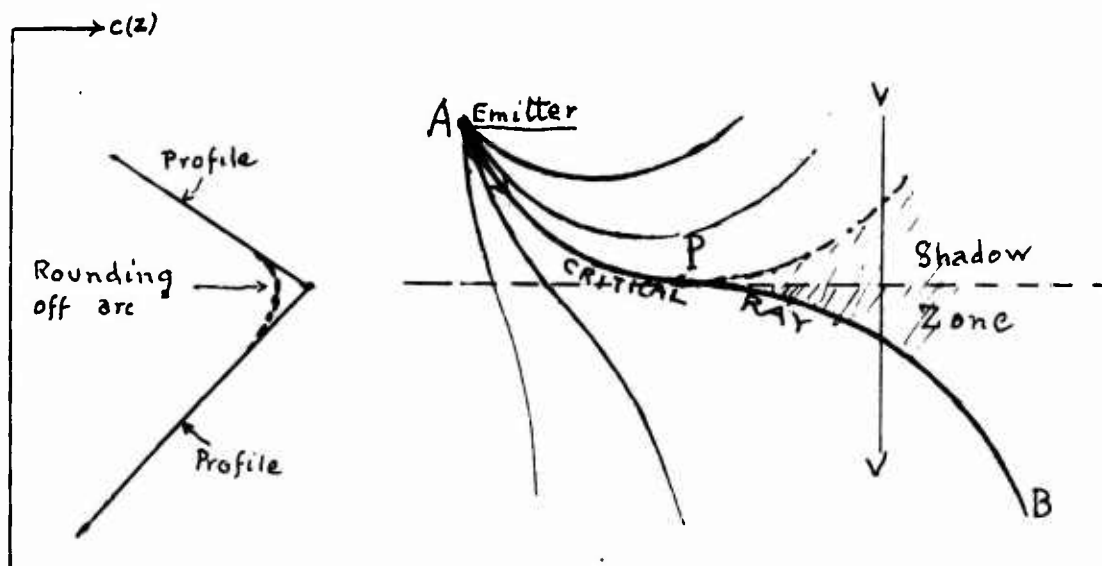


Figure A-III. Circular Construction of Critical Ray and Shadow Zone

Everything stated in the last paragraph is mathematically above reproach, being rigorously derived from Fermat's principle in the case of broken line profiles. Physically, on the other hand, we may have serious misgivings, and in fact ask two questions: First, are angular junction points such as  $(z_2, c(z_2))$  physically possible? The answer is easy to give: because of diffusion of salinity, heat conduction, and the mixing (turbulence) on all scales, the sound speed cannot fail to be a smooth function of position. Even if by some process it could have a discontinuity in its gradient at a junction point at one epoch  $t = 0$ , these physical processes would at once obliterate it when  $t > 0$ . Second, isn't the linear approximation "good enough" for practical purposes?

This second question has to be answered by mathematics, since it involves a precise structural and quantitative comparison of the

consequences of two different assumptions. To begin with, if  $c(z)$  and hence  $n(z)$  are smooth--i.e., continuous functions with continuous derivative of at least the first two orders--the differential equations (A.2) or (A.8) are valid throughout the medium. It then follows from the uniqueness theorem of their solutions that if two rays are tangent at any point, such as P, they must coincide throughout their whole length. Hence, the critical ray as constructed in Figure A-III does not exist in Nature. Furthermore, the theorems on differential equations show that as the direction of the tangent at A to the ray varies continuously, so does the depth  $z$  of its intersection with any vertical line such as VV, at all points where it cuts VV' without being tangent to it. Therefore, if the angularity in the profile is rounded off so as to become smooth, as shown dotted in Figure A-III, a ray through A and initially headed to pass through P will, as its initial direction varies continuously, sweep through the whole sequence of VV within the shadow zone. Therefore this construction of a zone in which no direct rays from A pass is invalid. More generally, only an intersection with an absorbing boundary or the appearance of caustics can prevent the possibility of joining A to any second point in the medium by a ray: these obstructions and these only can cause shadow zones. Finally, when the profile is rounded off as shown in Figure A-III, still having a maximum  $c$  at  $z_2$ , the differential equations show that one ray through A approaches the horizontal line  $z = z_2$  as asymptote as the range increases indefinitely. This whole situation appears in greater clarity in the graphical representation by the "surface of section" in phase space, given in Section 14.

How can "so little difference" in profiles as that between the angular and the rounded off one shown in Figure A-III make so great a difference in the structure of the whole system of rays? Suppose that the range in depth through which the two profiles differ shrinks to zero, in what manner, if any, will the rays in the rounded case approach those in the angular? The answer to the first question is that the differential equations (A.2) contain not only the profile function but its first derivative: the angular profile has no derivative at  $z = z_2$ ; the derivative of  $c$  exists and is constant at both sides of this point, but makes a finite jump as this point is crossed. Therefore the coefficients of the differential equations are not "only slightly different". Finally, the assumption that as the smoothed profile approaches the angular, the solutions of the former will approach those of the latter is erroneous: it would be correct if the approach of coefficients (derivatives of  $c$ ) were uniform; but this is not the case. The "lumping" of constants-- here the gradient of the sound speed --leads in this case not to useful approximations but to false conclusions. Shadow zones exist and are important; but their position and intensity must be computed correctly-- not a difficult process in view of the laminarity ( $c = c(z)$ ) which may be assumed in the short ranges where we wish to know them. See Appendix B.

## APPENDIX B

### THE CASE OF CONSTANT ACOUSTIC PROFILE

The most commonly used approximation to  $n = 1/c$  is to assume that it depends on depth only. This means that the "acoustic profile" ( $c$  plotted against depth) is the same at all the times and places in which the propagation is being studied. This situation, whether applied to a "flat earth" model or to a "spherical earth", has been called the laminar case --the properties of the sea being thought of as having a layered horizontal structure. We shall treat this case in the present section, using the "flat earth" model.

We shall in general use rectangular coordinates  $(x,y,z)$  with  $z$  measured vertically downward, the  $xy$ -plane parallel to the ocean surface. The ray equations are then those of (A.2), and the acoustic profile is given by

$$(B.1) \quad c = c(z), \quad n = 1/c = n(z)$$

In the present case our work will gain in simplicity if we introduce a system of cylindrical coordinates  $(r,\phi,z)$  based on the above rectangular axes:  $z$  is the same as before, while  $r$  is the horizontal range and  $\phi$  the azimuth (angle between the  $xz$ -plane and the plane determined by the point and the  $z$ -axis) so that  $x = r \cos \phi$  and  $y = r \sin \phi$ , and the element of arc is now given by the formula  $ds^2 = dr^2 + r^2 d\phi^2 + dz^2$ . With these expressions, Euler's equations take on the form

$$\begin{aligned}
 (B.2) \quad & \frac{d}{dt}(n^2 \frac{dr}{dt}) = \frac{1}{n} \frac{\partial n}{\partial r} + n^2 r \left(\frac{d\phi}{dt}\right)^2 \\
 & \frac{d}{dt}(n^2 r^2 \frac{d\phi}{dt}) = \frac{1}{n} \frac{\partial n}{\partial \phi} \\
 & \frac{d}{dt}(n^2 \frac{dz}{dt}) = \frac{1}{n} \frac{\partial n}{\partial z}
 \end{aligned}$$

We shall assume that we have taken the z-axis passing through the point of emission O. Since n is independent of  $\phi$ , the second equation in (B.2) has its second member zero, from which it follows that, along each ray,

$$n^2 r^2 \frac{d\phi}{dt} = \text{constant.}$$

Since at the initial point O of this ray  $r = 0$ , this constant is zero; in other words, along the ray,  $\phi = \text{constant}$ . But this means that the whole ray lies in the fixed vertical half-plane, bounded by the z-axis and characterized by the constant azimuth angle  $\phi$ .

Turning to the first equation in (B.2), the second member is seen to be zero: its first term, because n is independent of r; its second, because of the fact established in the last paragraph. Consequently

$$n^2 \frac{dr}{dt} = k, \text{ constant along the ray.}$$

Throughout this study, differentials along individual rays are denoted by  $\underline{d}$ , in contrast to differentials in which only the parameters specifying rays vary, denoted by  $\underline{\delta}$ . The last equation

then has a familiar geometrical interpretation; note that obvious substitutions lead to the equations

$$(B.3) \quad n \frac{2dr}{dt} = n \frac{dr}{cdt} = n \frac{dr}{ds} = n \sin \gamma = k,$$

where  $\gamma$  is the angle between the positively directed  $z$ -axis and the tangent to the ray at the point on it being considered. Our result, therefore, is<sup>\*</sup>

$$\text{Snell's Law:} \quad n(z) \sin \gamma = k$$

The constant  $k$  characterizes the ray (through  $O$  and in the plane of given fixed  $\phi$ ); we shall call  $k$  the "Snell constant" of the ray. One expression for  $k$  is in terms of the depth  $z = z_0$  of the point of emission  $O$ , ( $c(z_0) = c_0$ ) together with the initial direction at  $O$ ,  $\gamma = \theta$  (obviously the co-latitude introduced in Appendix A). Another useful expression is in terms of the value at a point (if one exists) where the ray has a horizontal tangent, so that  $\sin \gamma = 1$ ; at such a point the ray has either a highest or a lowest point (it "vertexes" or "bottoms"). Thus we may write Snell's law in the forms

$$(B.4) \quad n(z) \sin \gamma = k = n_0 \sin \theta = n(\max z) = n(\min z).$$

Snell's law may be re-written in the form of a first order differential equation connecting the two variables  $r$  and  $z$ . From one of the forms contained in (B.3), namely  $n dr = k ds$ , and from the

---

\* I.e., "Snell's law throughout the medium"; cf. the discussion in Appendix A.

expression for the arc-length along a ray (with  $d\phi = 0$ ),  $ds^2 = dr^2 + dz^2$ , we eliminate  $ds$ , thus obtaining

$$(B.5) \quad k^2 dz^2 = (n^2 - k^2) dr^2$$

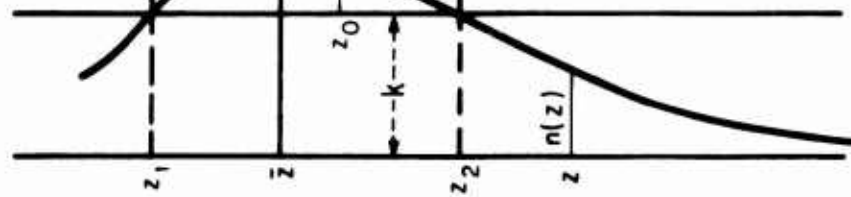
From this a number of conclusions can be drawn. First, that  $n^2 - k^2$  cannot be negative; and, since  $n$  and  $k$  are both non-negative, that  $n(z) \geq k$ . This shows that all the rays of a given value of the Snell constant  $k$  are confined to the horizontal layers bounded by the planes  $z = z_1$ , where  $z_1$  is determined by the equation  $n(z_1) = k$ . A simple case is shown in Figure B-Ia, in which the acoustic profile  $c(z)$  has a single minimum at  $z = \bar{z}$ , so that its reciprocal  $n(z)$  has a single maximum at this depth  $\bar{z}$ . For the value of  $k$  shown, there are two depths  $z_1$  and  $z_2$  for which  $n = k$ ; thus the ray having this Snell constant must lie between these. Figure B-Ib shows the ray in the case in which the vertical line distant  $k$  units from the  $z$ -axis cuts the graph of  $n(z)$  at non-zero angles; i.e., with derivatives  $n'(z_1) > 0$  and  $n'(z_2) < 0$ . The ray is the graph of the solution  $z = g(r)$  of (B.5). We shall return to the nature of this solution later.

The second fact emerging from (B.5) is that the solution can be obtained by quadratures (the calculation of an indefinite integral). For, on separating the variables (B.5) becomes

$$(B.6) \quad \pm \int \frac{k dz}{\sqrt{n^2(z) - k^2}} = r + \text{constant},$$



a. GRAPH OF INDEX  $n(z)$



b. RAY IN DUCT. SNELL CONSTANT  $k$

$z = g(x, k)$ ; INVERSE OF INTEGRAL (3.6)

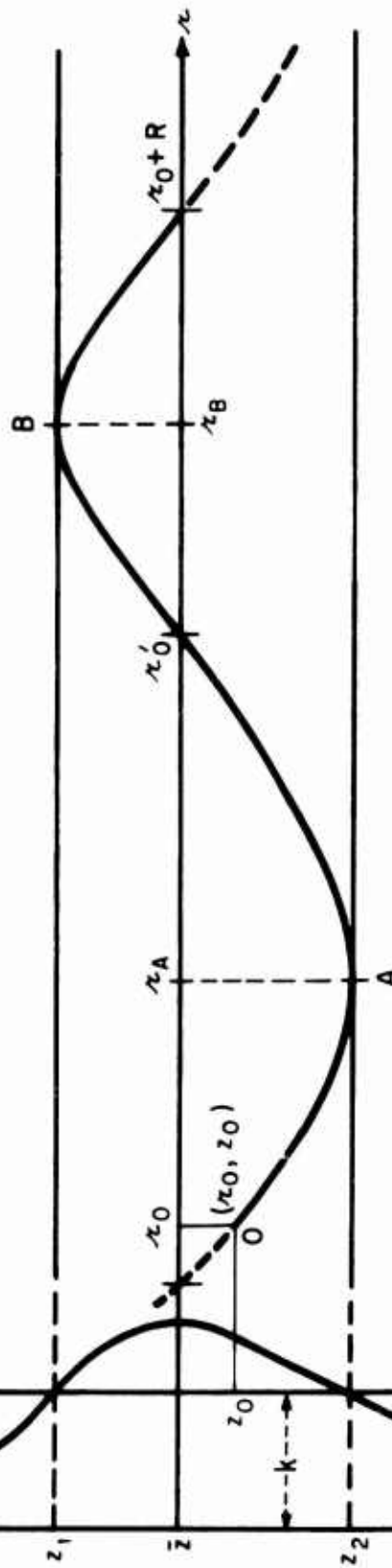


FIGURE B-1. TYPICAL RAY IN SIMPLE DUCT

where the (+) is taken when the ray slopes down (z and r both increasing) and the (-) when it slopes up (z decreases as r increases). Since  $n(z)$  is given as a graph or table, the integration may have to be done by numerical methods. However, when the profile (and hence  $n(z)$ ) is simple in character, it may sometimes be approximated by formulas making an explicit working out of the integral possible. But in any case, (B.6) defines r as a multiple valued function of z, having a single-valued inverse (graphed in Figure B-Ib).

$$(B.7) \quad z = g(r, k).$$

There are many other possibilities (not shown in the B-I figure) depending on the shape of the profile and the value of the given Snell constant: thus B-II (end) shows the asymptotic ray when c has a maximum (contrast it with A-III).

The last conclusion is that the solution of the ray equations (B.2)-- which we have carried out geometrically-- so that every ray is described by the equations  $\phi = \text{constant}$ , and z given by (B.7) (equivalent to (B.6))-- can be completed kinematically, i.e., so as to give our cylindrical coordinates (r,  $\phi$ , z) as functions of time, t, just as (A.4) gave the solution of (A.2) as such functions. For this purpose we have but to take from (B.3) the relation  $dr = kdt/n^2(z)$ , and use it to eliminate dr from (B.5), obtaining

$$(B.8) \quad \pm \int \frac{n^2(z) dz}{\sqrt{n^2(z) - k^2}} = t + \text{constant}$$

Since  $r$  and  $t$  increase or decrease together, the same rule of choice of signs applies to (B.8) as to (B.6).

We have therefore solved (B.2) for the three coordinates  $(r, \phi, z)$  of a point tracing a ray through 0 and in the initial direction  $(\theta, \phi)$ , by determining them as functions of  $(t, \theta, \phi)$  by means of the above formulas. We must now evaluate the spreading factor; in other words, we must calculate the area ratio of (A.7) in the present case.

We could of course return to the original rectangular axes, expressing  $(x, y, z)$  in terms of the present cylindrical coordinates; but this indirect method leads to unnecessarily complicated computations. It is better to calculate the vector product in terms of the vector components of the  $\theta$ -curves and the  $\phi$ -curves (cf. Figure A-I) along the three mutually perpendicular directions at  $(r, \phi, z)$  in which  $r, \phi$ , and  $z$  increase, one at a time, the other two remaining fixed.

Figure B-II shows these directions, with the components  $(\Delta r, r\Delta\phi, \Delta z)$  of a typical vector  $\Delta v$  (not shown to avoid complicating the figure).

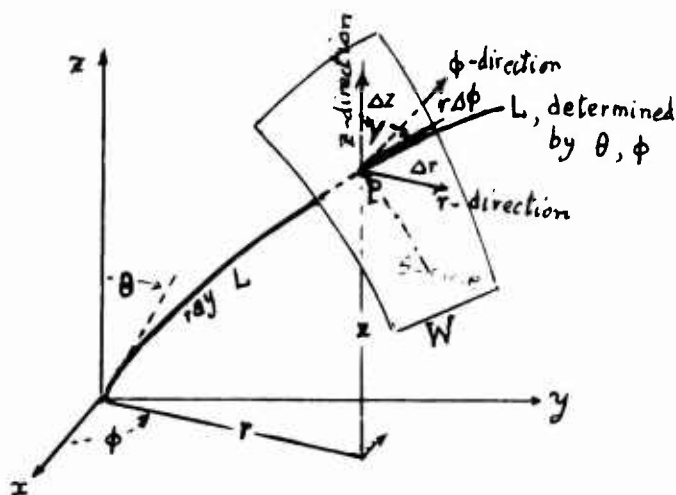


Figure B-II

Showing three mutually perpendicular directions:

- $r$  - direction
- $\phi$  - direction
- $z$  - direction

along which to resolve  $\phi$ -vector and  $\theta$ -vectors (not shown)

A simplifying feature of the present (laminar) case is that all figures (rays and wave fronts) have rotational symmetry about the z-axis: any ray through O going into a ray through O; and every wave front W is a surface of revolution about this axis. This is an immediate consequence of the rotational symmetry of the physical conditions, reflected in the absence of  $\phi$  from the equations (B.1), (B.3) - (B.7). Therefore the  $\phi$ -curves are circles in planes perpendicular to the z-axis, centered on the latter, and of radius r. Hence the elementary vector described by a point on W which has its parameter  $\theta$  fixed but its  $\phi$  increased by the amount  $\delta\phi$  is in the  $\phi$ -direction of Figure B-II and of length  $r\delta\phi$ . Its rectangular components in the  $(r, \phi, \theta)$ -directions are

$$\underline{\phi\text{-vector}}: \quad (0, r\delta\phi, 0)$$

To find the components of the corresponding  $\theta$ -vector, which represents the displacement of a point on W when  $\theta$  increases by the amount  $\delta\theta$  ( $\phi$  remaining fixed), we may confine ourselves to a half-plane through the z-axis, and observe at once that the component in the  $\phi$ -direction is zero. The components in the r- and z-directions are the differentials  $\delta r$  and  $\delta z$ , computed on the assumption that both  $\phi$  and t remain constant (the latter, because we are staying on a fixed wave front)-- but that the Snell "constant" k varies: it is constant along each ray, but varies from ray to ray, as we see from the formula derived from (B.4):

$$\delta k = \delta(n_o \sin\theta) = n_o \cos\theta \delta\theta$$

In the light of these facts, we have

$\theta$ -vector:  $(r, 0, z)$

Hence the vector product  $\vec{r}_\theta \times \vec{r}_\phi$  needed in (2.7) has the components in the  $(r, \phi, z)$ -directions given as follows:

$$\vec{r}_\theta \times \vec{r}_\phi : (-r \delta z \delta \phi, 0, r \delta r \delta \phi)$$

Therefore its absolute value is  $r \delta \phi (\delta z^2 + \delta r^2)^{1/2} = r \delta \phi \delta s$ , where  $\delta s$  is the differential of arc along the curve of intersection of the wave front  $W$  and the half-plane  $\phi = \text{constant}$ . In our figure of revolution, this expression is almost geometrically self-evident without calculating the vector product: our reason for doing this is to illustrate the much more general method.

To find  $\delta s$ , we note that in the half-plane through the  $z$ -axis, using the axes of  $r$  and  $z$  as horizontal and vertical rectangular axes, the slope of a ray is  $\text{ctn} \gamma = dz/dr$  (differentials being taken along the ray); whereas the slope of the  $\theta$ -curve is  $\delta z/\delta r$ . These directions being mutually perpendicular, we have  $\delta z/\delta r = -\tan \gamma$ . Eliminate  $\delta r$  and apply (B.3); we obtain

$$\delta s^2 = \delta r^2 + \delta z^2 = (\text{ctn}^2 \gamma + 1) \delta z^2 = \text{csc}^2 \gamma \delta z^2 = \left( \frac{n(z)}{k} \delta z \right)^2$$

The result is that  $\delta s = [n(z)/k] |\delta z|$ , the absolute value sign being required by the fact that  $\delta z$  may be positive or negative, while  $\delta s$  is essentially positive.

It remains to express  $\delta z$  in terms of  $\delta\theta$  (or, equivalently, of  $\delta k$ ). It is at this point that we must make the general relations given in (B.6) and (B.8) more precise. Let the ray represented in Figure B1b emanate from the point source 0 where  $r=0$ ,  $z=z_0$ ,  $t=0$ ; out to the first vertexing or bottoming, we have

$$(B.9) \quad r = \pm \int_{z_0}^z \frac{k dz'}{\sqrt{n^2 - k^2}}, \quad t = \pm \int_{z_0}^z \frac{n^2 dz'}{\sqrt{n^2 - k^2}}$$

the signs being chosen as explained earlier.

To find  $\delta z$  we differentiate the second formula, setting  $\delta t=0$  (the arc being on W) and allowing only  $k$  and  $z$  to vary. Solving the resulting equation for  $\delta z$  in terms of  $\delta k$  etc., and then substituting the value found into the earlier expression for  $\delta s$  in terms of  $\delta z$ , we obtain

$$(B.10) \quad \delta s = L(z, k) \delta k$$

where  $L(z, k)$  involves an integral and may be written in the two forms

$$(B.11) \quad L(z, k) = \frac{\pm 1}{n(z)} [n^2(z) - k^2]^{1/2} \int_{z_0}^z \frac{n^2(z) dz}{[n^2(z) - k^2]^{3/2}}$$

$$= \cos \gamma \int_{z_0}^z \frac{n^2(z) dz}{[n^2(z) - k^2]^{3/2}}.$$

The second form makes use of (B.4). The right-hand member is essentially positive between 0 and the point of first horizontal tangent

It follows that

$$(B.12) \quad \delta S = r \delta \phi \delta s = r L(z, k) \delta k \delta \phi$$

so that the required spreading factor is given by

$$(B.13) \quad \frac{\delta S}{\delta S_0} = \frac{r L(z, k) \text{ on } W}{r L(z, k) \text{ on } W_0}.$$

A particularly simple form occurs when the inner wave front  $W_0$  is taken so close to the emitter 0 that the changes in index  $n(z)$  in and on it can be ignored. Then  $W_0$  is a sphere centered at 0 as in Figure A-I; if its radius is denoted by  $a$  so that on the sphere

$$r = a, \quad z = z_0 = a \cos \theta, \quad n(z) = n_0$$

(B.11) shows that on  $W_0$ ,  $L(z, k) = a/n_0 \cos \theta$ , so that (B.12) becomes (using  $\delta k = \delta(n_0 \sin \theta) = n_0 \cos \theta \delta \theta$ )

$$(B.14) \quad \delta S_0 = \frac{a^2 \sin \theta}{n_0 \cos \theta} \delta k \delta \phi = a^2 \sin \theta \delta \theta \delta \phi$$

This is, of course, the usual element of area on a sphere in spherical coordinates. In spite of the fact that  $\cos \theta$  may be negative or vanish, the final expression is free from such ambiguities or singularities. Note that, by construing these steps as calculations of a limit as  $\theta \rightarrow \pi/2$ , everything is valid even when the initial tangent at 0 is

horizontal.

In applying (B.14) to the evaluation corresponding with (B.13), we must realize that the common angle  $\delta\theta$  rather than  $\delta k$  has been cancelled; hence on W we must write (B.12) in the form (using  $k=n_0 \sin\theta$ )

$$\delta S = n_0 r \cos\theta \cdot L(z,k) \cdot \delta\theta \delta\phi$$

We thus obtain, instead of (B.13), the spreading factor

$$\frac{\delta S}{\delta S_0} = \frac{n_0 r}{a^2} \cot\theta L(z,k)$$

which leads to the acoustic intensity ratio (cf.[1.1])

$$(B.15) \quad \frac{\text{Acoustic intensity at P}}{\text{Acoustic intensity at } P_0} = \frac{\delta S_0}{\delta S} = \frac{a^2 \tan\theta}{n_0 r L(z,k)} .$$

A more usual expression assumes that the acoustic intensity (flux of power per unit area) over the sphere  $W_0$  is constant and hence that the total rate of emission of energy by the source,  $E_0$ , is the spherical area  $4\pi a^2$  time the acoustic intensity. This replaces (B.15) by the expression

$$(B.16) \quad \text{acoustic intensity at P} = \frac{E_0 \tan\theta}{4\pi r n_0 L(z,k)} .$$

A concluding remark is in order concerning possible singularities in the integrand in (B.9). These occur for  $z$  such that



$n(z) = k$ , as at  $z_1$  and  $z_2$  in Figure B1; they correspond to horizontal tangency of the ray. As noted above, no difficulty or exception occurs when horizontality takes place at the emitter 0. It is the first point of horizontal tangency farther out on the ray that must be examined. At such a point  $z_i$ ,  $\sin \gamma = 1$ ,  $k = n(z_i)$ , and the integral in  $L(z, k)$  becomes infinite having the  $3/2$  power of  $z - z_1$  or  $z - z_2$  in the denominator of the integrand (see Fig. 5a). On the other hand,  $L(z, k)$  involves a factor that vanishes at these points: it is an "indeterminate form  $0 \times \infty$ ". We find it by l'Hospital's rule, by writing it as the integral divided by the reciprocal of its coefficient: the indeterminate form  $\frac{\infty}{\infty}$ , which is the limit of the ratio of the derivatives of the numerator and denominator. An easy calculation gives

$$\lim_{z \rightarrow z_2} L(z, k) = - \frac{1}{n'(z_2)}$$

$$\lim_{z \rightarrow z_1} L(z, k) = \frac{1}{n'(z_1)}$$

In the case illustrated in Figure B1, the two derivatives in the numerators are  $n'(z_1) > 0$  and  $n'(z_2) < 0$ , so that the values given by (B. 16) are positive and finite. Inserting the values in (B. 16) ( $i=1, 2$ ), it becomes

$$(B. 17) \quad \text{Acoustic intensity at } P = \frac{E_0 \tan \theta |n'(z_i)|}{4\pi r n_0} .$$

On the other hand if, e.g.,  $n'(z_2) = 0$ , as would occur at a local minimum of  $n(z)$ — a maximum of  $c(z)$ — (B.17) would suggest that the intensity at P is zero. While this happens to be true, the use of (B.17) in its proof is not directly applicable, since an inspection of (B.9) at such a point  $z_0$  by routine methods of the calculus shows that, as  $z \rightarrow z_2$ , both  $r$  and  $t$  increase without limit : there is at  $z=z_0$  a horizontal unstable ray that is approached as an asymptote by the ray under consideration (assumed not initially horizontal: it would otherwise coincide with the horizontal asymptote). There is in fact no first point of horizontal tangency to our ray. This assumes that all acoustic profiles occurring in Nature are smooth curves (have continuous first derivatives). All this has been discussed in Appendix A; the asymptotic ray is shown in Figure B-II at the end.

We have thus obtained explicit expressions for the rays emitted from 0 and the acoustic intensity along them, in the region so close to 0 that no multiple-path transmission occurs (we are leaving out reflections). This is usually of the order of 30 miles. Our explicit expressions involve two integrals containing the acoustic index  $n(z)$ . The practical question of numerical computation is discussed in another study.

What do the explicit analytic formulas obtained in this Appendix tell us about the difficulties at longer ranges (caustics, etc.), described in general qualitative terms in Section 2. They were illustrated,

in the two-dimensional case by Figures 2-I and 2-II.

On applying our formulas to the analysis of these cases, exactly the same difficulties become apparent as have been illustrated geometrically. Thus if  $W$  is examined as the locus of the equation  $t = s(x,y,z)$ , (here,  $S(r,z)$ ) for various increasing constant values of  $t$ , equations(B.9) show that  $S$  is the function resulting from the elimination of  $k$ . In other words, we must regard the first equation as determining  $k$  as a function of  $(r,z)$ ; the symbol  $k$  in the second equation is interpreted as the function thus obtained; the integrand becomes therefore a function of the three variable,  $(r,z,z')$ ,  $z'$  being the variable of integration. But the conditions that the first equation determine  $k$  as a regular single-valued function of  $(r,z)$  are only satisfied "within a convergence zone" of the source; beyond such ranges, the singularities exhibited geometrically show up analytically.

It may be remarked that the use of the eikonal equation  $|\nabla S|^2 = n^2$  gives rise to exactly the same results. Transforming the integrand in the second equation in (B.9) by writing

$$\sqrt{\frac{n^2}{n^2 - k^2}} = \sqrt{\frac{(n^2 - k^2) + k^2}{n^2 - k^2}} = \sqrt{\frac{k^2}{n^2 - k^2}} + \sqrt{n^2 - k^2}$$

and then applying the first equation, we obtain

$$t = kr + \int_{z_0}^z \sqrt{n^2 - k^2} dz' ,$$

which is precisely what would have been obtained if we had solved the eikonal equation by the method of separation of variables, and then applied the initial condition that at the point  $O$  ( $r = 0, z = z_0$ ),  $t = 0$  for all values of  $\phi$ . But the above expression, while simple in itself, still contains  $k$ : its elimination from the first equation in (B.9) so as to obtain  $t = S(r, z)$  involves all the difficulties of the former method.

We close with a schematic illustration of the asymptotic ray occurring at a smooth maximum of  $c$ , as contrasted with the "critical ray" at the angular maximum of Figure A-III. With the aid of formulas of the sort developed above, the spreading factor and hence the (lowered) acoustic intensity in its neighborhood may be expressed (and computed numerically): it is seen as a region of rarified ray density.

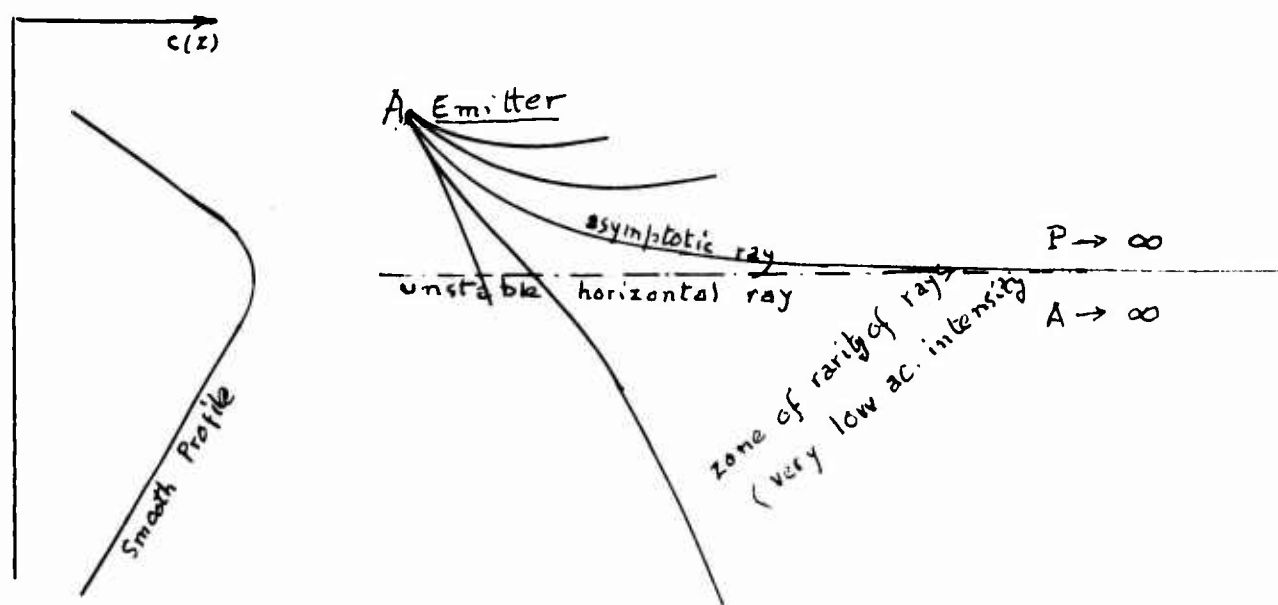


Figure B-III. Critical Asymptotic Ray

APPENDIX C  
INDIVIDUAL TRAVELLING WAVES

The purpose of this appendix is to examine more deeply than in Section 5 the physical and mathematical nature of the boundary conditions and the individual travelling wave, deferring to Appendix D families of such waves and their behavior with increasing circular frequency  $\omega$ . First, the reasons that allow us to assume the analyticity of the functions are set forth. Second, the mathematical nature of the boundary conditions is formulated. Third, a process of normalization is applied, completing the definition of the travelling waves and removing certain ambiguities that have plagued their extension to non-homogenous media. Finally, the theorem is proved showing that even on internal boundaries, where  $A = B = 0$ , so that the definition of  $S$  by (5.3, 4, 5) is not directly applicable,  $S$  can be extended by continuity to this locus and is single-valued in its neighborhood except for a finite jump of  $\pi/\omega$  as this boundary is crossed; since this drops out from the gradient  $\nabla S$ , it causes no indeterminacy in the direction of  $\vec{F}'$ , cf. (5.9), which latter approaches tangency to the internal boundary.

The physical medium in which the radiation occurs will be denoted by  $W$  (and assumed to be finite); and its material boundaries, by  $B$ . The basis of the treatment will be the pairs of real-valued functions, the elementary travelling wave pair  $(A, B)$  for the circular frequency  $\omega$ , and the general Helmholtz equation (4.2), which we will write as

$$(C.1) \quad \frac{1}{\rho} \nabla \cdot (\rho \phi) + k^2 \phi = 0, \quad k = \omega n = \omega/c;$$

we recall that  $\rho$  and  $k$  depend on position but not on time.

Analytic nature of the physical quantities. The given quantities  $\rho$  and  $c$  are "physical" functions and can of course be known only approximately by measurements, or calculations based on observations--certainly the accuracy stops before atomic dimensions are reached! Owing to diffusion and heat conduction they are continuous and have continuous gradients. It is often useful to use discontinuous approximations: e.g., a connected set of straight lines to represent an acoustic profile in studying layered media; or actual jump discontinuities across boundaries. But in such cases care is necessary to make sure that results so based are not essentially "artifacts" of the simplifying assumption.\* It is easy to show mathematically that such physical functions as  $\rho$  and  $c$  can be approximated arbitrarily closely by analytic functions: i.e. functions that can be expanded in convergent power series in the coordinates in the neighborhood of each point of the space of interest. In dealing with matter in bulk, as we are in this study, the properties deduced from the analytic approximation show themselves to be valid (at least in the limit) as the approximation gets better and better: this has in fact been the basis of centuries of success of the analytic assumption. The same is not necessarily true of the other (discontinuous) forms of approximation mentioned above.

The boundaries. The detection boundary is but a mathematically constructed surface enclosing the detection equipment, which, in the present context, exerts no physical effect on the radiation; it will therefore be ignored henceforth.

---

\* As in the case of the "critical ray" in the conventional construction of shadow zones discussed at the case of Appendix A.

The emission boundary  $E$ , on the other hand, corresponds to an essential feature of the field, since it contains the physical agents producing it: outside of  $E$  the wave equation (3.1) and (C.1) are valid; inside it they do not always hold. We shall assume that the emitter is relatively small, in the sense that  $E$  can be so drawn that inside it the two "quiet" functions  $\rho$  and  $c$  are essentially constant. Since this continues to be true when  $E$  is replaced by  $E'$  sufficiently close to  $E$ --e.g., obtained by slightly deforming it--we see that these two properties do not completely determine the emission boundary. This flexibility may be used to give simple shapes to  $E$ ; thus in our study in Sections 13, 14, 15 and 16 of emission with azimuthal symmetry, we have chosen  $E$  to be a vertical cylinder of revolution; in other applications, a spherical  $E$  is more useful. In the general case,  $E$  can always be chosen to be an analytic surface (i.e., the locus of equations imposed on analytic functions). Finally, the values of the wave functions  $A$ ,  $B$ ,  $\phi$ , etc., can be assumed to satisfy analytic conditions on this analytic surface, for the reasons stated above. It will follow that all the derivatives will be analytic also. This possibility will have consequences that will be examined later. But the essential property of the emission boundary is that through it passes the power emitted, at each frequency at which this takes place. The boundary conditions should therefore give the total mean flux across  $E$ :

$$(C.2) \quad \iint_E \vec{F}_N' \cdot d\vec{G} = \iint_E (\rho\omega/2) \left[ A \frac{\partial B}{\partial N} - B \frac{\partial A}{\partial N} \right] d\mathcal{G} = F_\omega(E)$$

Here  $\vec{N}$  denotes the unit vector normal to the boundary, oriented outward from  $W$ , and  $\partial/\partial N$  is the normal derivative--the normal component of the gradient along  $\vec{N}$ .

Turning, lastly, to the physical boundaries,  $S$ , we observe that they act as interfaces between pairs of media: water and air at the ocean's surface; water and quasi-solid matter at its bottom and borders. Across every such bounding surface, two conditions must be fulfilled: the normal components of the velocities of each medium must coincide; and so must the normal pressures. In terms of the velocity potentials  $\psi$  and  $\psi_1$  (the latter describing the bounding medium), these conditions are expressed by the equations

$$(C.3) \quad \frac{\partial \psi}{\partial N} = \frac{\partial \psi_1}{\partial N}, \quad \rho \frac{\partial \psi}{\partial t} = \rho_1 \frac{\partial \psi_1}{\partial t}.$$

The first equation results from the defining property of a velocity potential; the second, from the fact proved in theoretical acoustics, that  $-\rho \partial \psi / \partial t$  is, to the linear approximation used, equal to the fluid pressure. We are using the subscript, 1, for all quantities in the bounding medium. [Note that many authors employ  $-\psi$  instead of our  $\psi$ , with corresponding reversals of signs for velocity and pressure]. When the bounding medium is more complicated and requires stress tensors or viscosity coefficients to describe its motion, the conditions are expressed by more complicated equations than the above (see, e.g., Tolstoy and Clay, Chapter 2).<sup>1</sup> But in all cases, the conditions are linear. This means that if



(C.3) are verified by the two pairs,  $(\psi, \psi_1)$  and  $(\psi', \psi'_1)$ , then they will also be verified by  $(a\psi + b\psi', a\psi_1 + b\psi'_1)$ , where  $\underline{a}$  and  $\underline{b}$  are any two constants.

It is obvious that on the contrary the conditions on the emission boundary  $E$  are not linear: merely multiplying our functions  $(A, B)$  by a constant will multiply the mean power flux throughout  $\mathcal{W}$ , and across  $E$  in particular, by the square of this constant--changing the strength of the source by that factor.

There is, however, a method of forming linear combinations of solutions  $(A, B)$  which does not affect any mean flux vector  $\vec{F}$ , and therefore continues to satisfy (C.1), (C.2), (C.3); it is the unimodular transformation  $(A, B) \rightarrow (A', B')$ :

$$(C.4) \quad \begin{aligned} A' &= \alpha A + \beta B \\ B' &= \gamma A + \delta B \end{aligned} \quad \Delta = \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = 1.$$

A simple computation shows that

$$A' \nabla B' - B' \nabla A' = (A \nabla B - B \nabla A) \Delta.$$

Hence, for the preservation of the mean flux vector field, it is necessary and sufficient that  $\Delta = 1$ . Note that to take  $\Delta = -1$  (a reflection) would merely reverse the sense of each vector. Note also that if instead of  $A, B$  we had used the complex wave function  $C = A + iB$  of (5.12), (5.13), etc., and similarly for  $C'$ , our formula (C.4) would determine  $C'$  as a

linear function of  $C$ , but an analytic one (satisfying the Cauchy-Riemann differential equations) if and only if  $\delta = \alpha$  and  $\gamma = -\beta$ ; this condition leads to  $\Delta = \alpha^2 + \beta^2 = 1$ , and therefore that (C.4) is simply a rotation of axes in the  $uv$ -plane of the variables  $u = A$ ,  $u = B$ : it amounts to adding a constant "phase" to  $\theta$  or  $S$  of Section 5. This special case of (C.4) is the necessary and sufficient condition that the mean energy density  $E'$  corresponding with  $(A, B)$  be left unchanged, as a straightforward calculation based on (4.3) will show. Other special cases of (C.4) are the "principal axis transformation", with  $\delta = 1/\alpha$  and  $\beta = \gamma = 0$ ; and the "shear", with  $\alpha = \delta = 1$  and  $\beta = 0$  (or with  $\gamma = 0$ ); it is easy to show that every unimodular transformation (C.4) is the "product" of these three special cases (i.e., the result of their successive performance).

We shall use the flexibility offered by (C.4) to remove a notorious ambiguity in the concept of "travelling wave". This ambiguity is mentioned by Brekhovskikh [Chapter III, Section 17, Section 6]<sup>1</sup>, who emphasizes that the splitting of the total radiation field near a reflecting barrier into the sum of an incident and a reflected wave is unambiguous only in the case of homogenous media. He illustrates the ambiguity by an example due to Schelkunoff: the complex valued function on the real  $z$ -axis is

$$C(z) = \cos kz + e^{ikz},$$

giving in our notation, by

$$A = (1 + \epsilon) \cos kz, \quad B = \epsilon \sin kz .$$

When  $\epsilon \ll 1$ , the function describes an essentially standing wave. But we can write

$$C = R(z) e^{i\theta(z)}$$

$$\text{with } R = [(1 + \epsilon)^2 \cos^2 kz + \epsilon^2 \sin^2 kz]^{1/2}$$

$$\theta(z) = \tan^{-1} \left( \frac{\epsilon}{1 + \epsilon} \tan kz \right).$$

And this expression takes the form of a travelling wave.

A general way of characterizing the ambiguity is to say that, however we may write an expression for a travelling wave, we may usually add standing waves of the same frequency, with a zero mean flux through the boundaries, so that we have many radiation fields with the same mean flux of energy.

Intuitively, the essence of the notion of "travelling wave" is the displacement of radiant energy and not its accumulation in "stagnant pools". To express this idea precisely, we seek--for a given pair (A,B)--that linear combination (A', B') which, while giving the same vector flux field, has the least total mean energy in the field. Evidently we must choose that transformation of type (C.4) which minimizes the quantity

$$\text{total mean energy in } \omega, \quad E(\omega) = \iiint_{\omega} E \, dx dy dz.$$

Writing, for brevity,  $u$  and  $v$  being any real functions of  $(x,y,z)$ ,

$$Q(u,v) = \iiint_W \rho(\nabla u \cdot \nabla v + \omega^2 n^2 uv) dx dy dz,$$

we see that, according to (4.3),

$$4E(W) = Q(A, A) + Q(B, B).$$

Suppose that  $(A, B)$  is, of all  $(A', B')$  given by (C.4), the one which minimizes the  $E(W)$ . Then the axial case of this transformation can only increase it. But this gives

$$Q(A', A') + Q(B', B') = \alpha^2 Q(A, A) + Q(B, B)/\alpha^2$$

and if the minimum of this must occur for the identical transformation--i.e., for  $\alpha = 1$ --we must have  $Q(A, A) = Q(B, B)$ . Assuming this relation, we apply the shear  $(A' = A, B' = \gamma A + B)$ , obtaining

$$Q(A', A') + Q(B', B') = (1 + \gamma^2) Q(A, A) + 2\gamma Q(A, B) + Q(B, B)$$

In order that the left hand member be a minimum when  $\gamma = 0$  (the identity) we must have  $Q(A, B) = 0$ .

Conversely, it is easily shown that when  $(A, B)$  satisfy the equations

$$(C.5) \quad Q(A, A) = Q(B, B) ; \quad Q(A, B) = 0,$$

then the total mean energy  $E(W)$  is a minimum.

Definition. The pair  $(A, B)$  of solutions of (C.1), (C.2), (C.3), for which (C.5) is true is called the travelling wave corresponding to this pair of solutions.

On applying this criterion to the  $(A, B)$  of Schelkunoff's example, given above, we determine the coefficients in (C.4) so that the new (C.5) is satisfied. We find first that the integrand in  $Q(A, B) = 0$ . To complete the determination, we look for the  $\alpha$  for which, after the axial transformation  $A' = \alpha A, B' = B/\alpha, Q(A', A') = Q(B, B')$ . We obtain  $\alpha = [\epsilon/(1 + \epsilon)]^{1/2}$ , so that

$$A' = \sqrt{\epsilon(1 + \epsilon)} \cos kz, B' = \sqrt{\epsilon(1 + \epsilon)} \sin kz$$

Whence,  $C' = \sqrt{\epsilon(1 + \epsilon)} e^{ikz}$ , determining a pure travelling wave without ambiguity.

Another example is the application to waves along a homogenous stretched string, along the  $x$  axis. The general solution of the wave equation is  $f(x - ct) + g(x + ct)$ . On applying the criterion for minimum energy for a given energy flux, we find that either  $g = 0$  (forward wave), or  $f = 0$  (backward wave). All other familiar cases such as plane, cylindrical, and spherical waves are divested of all ambiguity by applying this minimum energy criterion.

The branch-locus theorem. For the reasons set forth earlier in this appendix, we may assume that the quantities  $\rho$  and  $n = 1/c$  appearing in (C.1) are analytic throughout  $W$  and its boundaries  $B$  and  $R$ , that the latter are made up of analytic surfaces, and that all data and conditions on these boundaries are analytic. Combining these properties with the fact that (C.1) is an elliptic equation, we may assume, in view of the extensions of a classical theorem (Courrant-Holbert, I, Chap. VI.18)<sup>3</sup> that the solutions  $A$ ,  $B$ , etc. are analytic functions throughout  $W$ . On the basis of this fact we shall prove the branch-locus theorem:

THEOREM. In the travelling wave  $(A, B)$  satisfying (C.1), any regular 2-dimensional part  $\Gamma'$  of the singular locus  $\Gamma$  of  $A = B = 0$  causes no indeterminacy in direction of  $\vec{F}'$ --which approaches tangency  $\Gamma'$ --and causes only a finite jump in  $S$  by the constant  $\pi/\omega$  in crossing  $\Gamma'$ .

Being defined by analytic equations,  $\Gamma$  is composed of smooth analytic pieces, possibly joined together at conical points or singular curves. Let  $P_0: (x_0, y_0, z_0)$  be a point on a regular piece  $\Gamma'$  of  $\Gamma$  (i.e., not on such lower-dimensional singularities). In a neighborhood  $N_0$  of  $P_0$ , the corresponding piece of  $\Gamma$  can be represented as the locus of an equation of the form  $f(x, y, z) = 0$ , where  $f$  is analytic and  $\nabla f \neq 0$  throughout  $N_0$ . Then by a classical theorem in analytic functions of several variables,\* since the vanishing of  $f$  implies that of both  $A$  and  $B$ , we must have throughout a sufficiently small  $N_0$ ,

---

\* The Weierstrass Theorem.

$$A = af^p, \quad B = bf^q,$$

where  $p$  and  $q$  are positive integers, and  $a$  and  $b$  are analytic and never both zero within  $N_0$ .

Suppose that  $p > 1$ . Then

$$\nabla A = f^{p-1} (f \nabla a + pa \nabla f).$$

Therefore both  $A = 0$  and  $\nabla A = 0$  on the 2-dimensional piece of  $\Gamma$  for which  $f = 0$ . Then by the Cauchy-Kovaleski uniqueness theorem for elliptic partial differential equations in the analytic case, \* we should have  $A = 0$  throughout  $N_0$ --and hence, by analytic continuation, throughout  $W$ --contrary to its assumed properties. Therefore  $p = 1$ ; and, similarly,  $q = 1$ ; thus we have in  $N_0$

$$A = af, \quad B = bf, \quad R = \sqrt{A^2 + B^2} = |f| \sqrt{a^2 + b^2} = |f|r$$

Now the defining equation (5.5) becomes, in  $N_0 - \Gamma'$ ,  
 $a = \sigma(f)r \cos \omega S$ ,  $b = \sigma(f)r \sin \omega S$ , where  $\sigma(f) = |f|/f = \pm 1$ . Since on crossing  $\Gamma'$ ,  $f$  and therefore  $\sigma(f)$  reverses sign, it follows that  $\theta = \omega S$  jumps by  $\pi$ , so that  $S$  behaves as stated in the theorem.

This shows that the field of unit directional vectors  $U = F_{11}^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{r}_1}$ , which is uniquely defined by (4.4) (or (5.9) or (5.13)) everywhere except on  $\Gamma$ , is also defined on all 2-dimensional pieces  $\Gamma'$  of  $\Gamma$ . Furthermore, this field is "smooth" (continuously differentiable to any required order) at all such pieces, since its defining data are analytic.

---

\* Hadamard, "Lectures on Cauchy's Problem..."<sup>3</sup>

But this in itself does not show that the field is tangent to each  $\Gamma'$ :  
in this sense, that  $\Gamma'$  is an internal boundary.

To establish the tangency, we observe that at any point of  
 $N_0 - \Gamma'$  we have

$$\omega \nabla S = \nabla \theta = (A \nabla B - B \nabla A)/R^2 = (a \nabla b - b \nabla a)/r^2$$

Then we start from the equation of continuity of power flux,

$$(C.3) \quad \nabla \cdot \rho(A \nabla B - B \nabla A) = 0$$

which is easily derived from (C.1) (write  $\phi = A$  and then  $\phi = B$ ; multiply  
the first by  $-A$  and the second by  $B$ ; add; then manipulate the derivatives  
in an obvious way (this is an alternative proof  $\nabla \cdot \vec{F} = 0$ )). By replacing  
 $A$  and  $B$  by  $af$  and  $bf$  we obtain the equation

$$(a \nabla b - b \nabla a) \cdot 2f \nabla f + f^2 \nabla \cdot (a \nabla b - b \nabla f) = 0$$

Apply this to any point of  $N_0$  other than on  $\Gamma$ , divide by  $f$ , and let  
 $(x, y, z)$  approach any point on  $\Gamma'$ : this shows that  $\nabla S \cdot \nabla f = 0$  on this  
surface: since, as stated above,  $\nabla f \neq 0$  on  $\Gamma'$ ,  $\nabla f/|\nabla f|$  is the normal to  
 $\Gamma'$ ; and our statement of the tangency of the power flow lines to  $\Gamma'$  is  
proved.



## APPENDIX D

### WAVES, RAYS AND POWER FLOW LINES

The purpose of this appendix is to supply the mathematical details that underlie the more descriptive discussion in the text of the families  $\{\phi_\omega\}$  and  $\{\psi_\omega\}$  of elementary travelling waves, their power flow lines ( $\Lambda$ -lines) and rays (L-lines) from the point of view of curvature and torsion. Appendix G will show how, by using Hadamard's elementary solution of the wave equation (corresponding physically to an instantaneous pulse from a point), a mathematical process (intuitively, a sort of harmonic modulation of a train of pulses) will lead to a Fermat family in the broadened sense defined at the close of Section 6. This goes beyond the conventional treatments by enabling us to handle the cases of multiple-valued phase front functions, caustics, and multiple-path transmission. We start by giving the six examples referenced in Section 6, and conclude with the needed differential geometry.

The first example is that of the square homogenous drumhead, or membrane clamped at its edge. It is discussed, e.g., in Morse and Ingard,<sup>1</sup> Chapter 5.2, where it is used to illustrate the case of "eigenvalue degeneracy" (p. 206)--the only one in which a net mean displacement of energy is possible. Taking the units in such a way that  $\rho = c = 1$  and the side of the square is  $\pi$  units of length, the real solutions of the Helmholtz equation for a given frequency  $\omega$  come in pairs

$$A = \sin nx \sin my, \quad B = \sin mx \sin ny$$

with  $\omega^2 = n^2 + m^2$ . Degeneracy (distinct pairs) occur if and only if the positive integers  $(m, n)$  are different. Consider the case in which

$m = 1$  and  $n$  is allowed to increase without limit, so that  $\omega^2 = 1 + n^2$  does likewise. The locus of  $A = 0$  are the equi-spaced vertical lines for which  $\sin nx = 0$ , i.e., of equations  $x = k\pi/n$ ,  $k = 0, 1, \dots, n$ . The locus of  $B = 0$  are the horizontal lines,  $y = j\pi/n$ ,  $j = 0, 1, \dots, n$ . Therefore the singular locus of the  $S$  defined in Section 5 and Appendix C consists of the lattice points of coordinates  $(k\pi/n, j\pi/n)$ . These are regularly spaced throughout the square, and with increasing frequency ( $n \rightarrow \infty$ ) crowd everywhere densely throughout this region. There is circulation of power around each singular point, as an application of (5.9) shows: the  $\Lambda$ -lines become tighter and tighter around these crowding points and are evidently without relationship whatsoever to the  $L$ -lines which are all straight, since  $\rho$  and  $c$  are constant.

The two further examples also involve the homogenous membranes, in the first case clamped about a circle; it is discussed at p. 209 of the preceding reference. Separation of variables in Helmholtz' equation gives the classical solution [cf. (5.2.22) of preceding reference] which, in the polar coordinates  $(r, \theta)$ , we may write as the pair

$$A = J_p(\omega r) \cos p\theta, \quad B = J_p(\omega r) \sin p\theta,$$

where  $J_p$  is the Bessel function of positive integral order  $p$ . When the units are chosen so that  $c \neq 1$ ,  $\omega$  is to be replaced by  $\omega/c$  in the above. The singular locus  $A = B = 0$  is the locus at which  $J_p(\omega r) = 0$ : formed by concentric circles of radii  $r = B_{pq}/\omega$ , where  $B_{pq}$  is the  $q$ 'th zero of  $J_p(x)$  [cf. (5.2.23) of reference]; and crowding increasingly together as  $\omega$  increases indefinitely. On the other hand,  $S$  has only

one branching singular locus: the origin,  $r = 0$ . Everywhere else  $\theta$ , and with it  $S = \theta/\omega$ , are determined by (5.3) or (5.4), which show that  $\theta = p\theta$  and  $S = (p/\omega)\theta$ . This illustrates the basic theorem proved at the end of Appendix C, that although the singular locus may be only one dimension lower than the radiation space, the directional ambiguity part is always two dimensions lower. Also, since the  $\Lambda$ -lines are all concentric circles, the circular parts of the singular locus are internal boundaries.

But the important point illustrated by this second example is that the  $\Lambda$ -lines, which are circles, cannot be replaced by the L-lines, which are all straight.

The third example is also based on a homogenous membrane stretched over the plane, but now provided with a point source of radiation, situated at the origin of the polar coordinate system. The separation of variables applies as before, but the appropriate solutions of Bessel's equation with the central singularity represented by the source is the pair of functions  $J_p$  and  $N_p$  (the Bessel and Neumann functions). These are often combined into the two complex valued Hankel functions:

$$H_p^{(1)} = J_p + iN_p, H_p^{(2)} = J_p - iN_p.$$

Thus we have, in terms of real functions, four solutions (two, when  $p = 0$ ), products of these (in which the variable is replaced by  $\omega r/c$ ) by  $\cos p\theta$  and  $\sin p\theta$ . The only ones relevant to emission are

the ones containing both  $J_p$  and  $N_p$ , for example

$$A = J_p(\omega r/c) \cos p\theta, \quad B = N_p(\omega r/c) \cos p\theta$$

In the simplest case,  $p = 0$ , and our solutions are  $J_0(\omega r/c)$ ,  $N_0(\omega r/c)$ . Since these are independent of  $\phi$ , the function  $S = \tan^{-1}(N_0/J_0)/\omega$  depends only on  $r$ , and hence the  $\Lambda$ -curves are radial straight lines: the one case in which they coincide with the L-curves (rays). When  $p > 0$ , various elaborate  $\Lambda$ -curves are obtained, quite unrelated to the L-curves, which are always straight. In this case again the singular branch locus is a single fixed point, the origin; although the internal boundaries  $A = B = 0$  are 1-dimensional, crowding closer and closer together with increasing  $\omega$ --but non-existent when  $p = 0$ .

In these examples, the functions  $A$  and  $B$  were not velocity potentials but displacements. The mathematical results are the same as for velocity potential in the acoustic case, the medium being a compressible fluid confined between two parallel absolutely rigid plates.

The fourth example is a 3-dimensional acoustic one, again as in the third example, a point source of emission in a homogenous medium. It is treated at length in Morse and Ingard,<sup>1</sup> Chapter 7.1. Introducing spherical coordinates  $(r, \theta, \phi)$ , the method of separation of variables in Helmholtz' equation leads to solutions which are products of functions of  $(\theta, \phi)$  which are surface spherical harmonics, by functions of  $r$ , which are powers of  $1/r$  times  $e^{i\omega r/c}$ --equivalently, sines and

cosines of  $\omega r/c$ . The simplest case is the spherically symmetrical point source, in which the solutions depend only on  $r$ ; we have then, for example,

$$A = \frac{1}{r} \cos (\omega r/c), \quad B = \frac{1}{r} \sin (\omega r/c)$$

Clearly  $S = r/c$  and hence  $|\nabla S|^2 = 1/c^2$ : the eikonal equation. The  $\Lambda$ -lines are radial straight lines and coincide with the  $L$ -lines through the origin. Finally, the singular locus  $A = B = 0$  is non-existent; but the origin is a singularity of  $A$ . In the higher order cases, the solutions involve angular dependence, through the surface spherical harmonics; they correspond to multipole emission; the  $\Lambda$ -lines are curved, and only tend to straighten out with increasing  $r$ : in the "far field" the ray treatment becomes a satisfactory approximation.

The fifth example is the case of reflection and refraction of a plane wave at the interface ( $y = 0$ ) between homogenous media, the incident ray impinging from the lower medium of constant  $\rho, c$  ( $y < 0$ ),

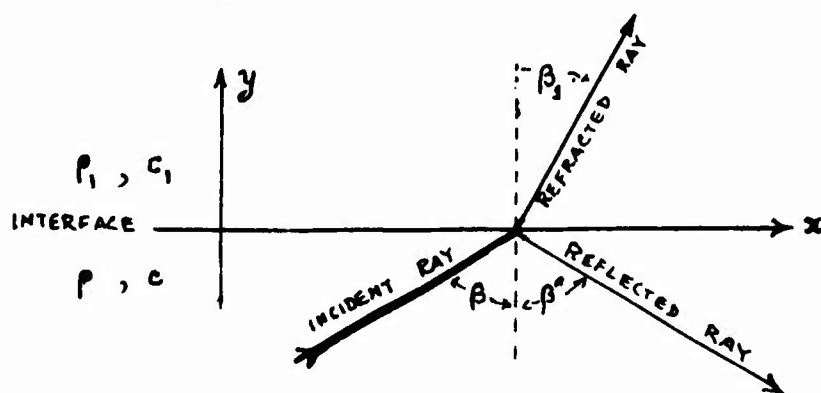


FIGURE D.I. REFLECTION AND REFRACTION

reflected back, as well as refracted into the upper medium of constant  $\rho_1, c_1$  ( $c_1 < c$ ); Figure D.I.

If we set the three waves equal to

$$\begin{aligned}\phi_{\text{inc.}} &= a \exp [i\omega (x \sin \beta + y \cos \beta)/c] \\ \phi_{\text{refl.}} &= b \exp [i\omega (x \sin \beta' - y \cos \beta')/c] \\ \phi_{\text{refr.}} &= r \exp [i\omega (x \sin \beta_1 + y \cos \beta_1)/c_1],\end{aligned}$$

then the total radiation below the interface is  $\phi = \phi_{\text{inc.}} + \phi_{\text{refl.}}$ ; this must be matched to that above the interface,  $\phi_1 = \phi_{\text{refr.}}$ , by applying the two interface conditions of Appendix C, which in the present case are that when  $y = 0$ ,  $\partial\phi/\partial y = \partial\phi_1/\partial y$  and  $\rho\phi = \rho_1\phi_1$ . These yield all the classical results, such as  $\beta' = \beta$ ,  $\sin \beta/c = \sin \beta_1/c_1$  (Snell's interface law) as well as the two further equations

$$(a-b) \cos \beta/c = r \cos \beta_1/c_1 ; (a+b) \rho = r\rho_1.$$

From these the classical properties of the coefficients of reflection and refraction are obtained. See Morse & Ingard<sup>1</sup> Chapter 6.3, Brekhovskikh,<sup>1</sup> Chapters II and III.

What is important to us is the computation of the energy and power flux densities. We find from the formulas of Section 5, in upper medium

$$\begin{aligned}F_x^1 &= \left( \rho_1 \omega^2 r^2 / c_1 \right) \sin \beta_1, \quad F_y^1 = \left( \rho_1 \omega^2 r^2 / c_1 \right) \cos \beta_1 \\ E^1 &= \rho_1 \omega^2 r^2 / c_1^2\end{aligned}$$

just as we should expect; but in the lower medium

$$F_x = (\rho\omega^2/c) [(a^2 + b^2) \sin \beta + 2 ab I(\omega)]$$

$$F_y = [\rho\omega^2 (a^2 - b^2)/c] \cos \beta$$

$$E = (\rho\omega^2/2c^2) [a^2 + b^2 + 2 ab \sin^2 \beta I(\omega)]$$

where  $I(\omega) = \cos(2\omega y \cos \beta / c)$  may be described as the "interference term". As  $\omega \rightarrow \infty$  it merely oscillates between  $\pm 1$ , and thus the  $\Lambda$ -line direction  $F_x/F_y$  approaches no limit -- nor does the ration  $F_x:F_y:E$ .

It is noted, however, that the power flux across a vertical element of length  $\Delta y > 0$  involves the integral

$$\int_{y_1}^{y_1 + \Delta y} \cos(2\omega y \cos \beta / c) dy = \frac{c}{2\omega \cos \beta} \Delta \sin(2\omega y \cos \beta / c);$$

the  $\Delta \sin$  denoting the difference between the values at  $y_1$  and  $y_1 + \Delta y$ . Since this is a term never exceeding  $c/\cos \beta$ , divided by  $\omega$ , we see that it approaches zero as  $\omega \rightarrow \infty$ ; and this, for however small a positive  $\Delta y$ . For the same reason, the interference contribution to the total energy contained in an arbitrary area  $\Delta x \Delta y$  vanishes in comparison with the fixed part as  $\omega \rightarrow \infty$ . In this sense -- the "weak convergence" noted at the end of Section 6 and to re-appear at the end of the later Appendix G -- we can say that the ratios  $F_x : F_y : E$  approach those of the leading terms, namely

$$(a^2 + b^2) \sin \beta : (a^2 - b^2) \cos \beta : (a^2 + b^2)/2c$$

Hence the lower family is, in this "weak" but physically meaningful sense, an assymptatic family. It is not a Fermat family by the second definition of Section 6, since the energy speed  $F/E$  is never equal to the phase speed  $c$ .

Our sixth is a counter-example: while it has no physical significance as corresponding to an idealized experiment or mechanically realizable radiation, it is given for the following logical reason: Many authors (including the present one) have supposed that an acoustic wave which could be put into the form  $\psi = \phi_{\omega} e^{-i\omega t}$ , with  $\phi_{\omega} = R_{\omega} e^{i\omega S_{\omega}}$ , and in which  $R_{\omega}$  and  $S_{\omega}$  and their derivatives remained small as functions of  $\omega$ , would have their power flow more and more nearly along a fixed set of rays, so that ray tracing could be used for transmission loss calculations. This condition is evidently a necessary one (at least when sufficient detail of statement is supplied). The purpose of the present counter-example is to show that it is not sufficient. We consider the case of a homogenous medium ( $\rho$ ,  $c$ , constant) and the plane wave

$$\phi_{\omega} = \exp[i\omega(x \cos \omega + y \sin \omega)/c].$$

This satisfies the Helmholtz equation  $\nabla^2 \phi + \frac{\omega^2}{c^2} \phi = 0$ , and, for each fixed  $\omega$ , represents a plane wave in the direction (the normal to  $S = x \cos \omega + y \sin \omega$ ) making the angle  $\omega$  with the x-axis. Evidently  $R (= 1)$  and  $S$  have the properties of smallness (including all their derivatives) which are supposed sufficient conditions for the "increasingly accurate application of geometrical optics at high frequency" ("Try this out on your Victrola!").

We conclude this discussion of special cases by pointing out the fallacy of certain treatments of "layered" or "laminar" media -- i.e., in which the physical quantities  $\rho$  and  $c$  depend on depth,  $z$ , only. As explained in Appendix C, the physical causes of their variation guarantee that they are smooth functions of the position variables -- functions which may always be taken as analytic. On the other hand, certain modern authors



(e.g., L. M. Brekhovskikh<sup>1</sup>) use an approximation in which the layered medium is replaced by a "finely layered medium" composed of a large number of thin homogenous parallel "plates", separated by plane interfaces at which conditions similar to those of our fifth example are applied.

Given the above replacement of the smoothly layered medium by these parallel plates, the mathematical treatment is perfectly clear. But the assumption that the results in the second case approximate as closely as needed to those of the first must be rejected -- both on physical and on mathematical grounds. Physically, the plates would give rise to a large number of waves reflected back from the interfaces. As these increase in number, the waves' individual intensities decrease, but increase in number: no such effect has been observed in perfectly clear media. Mathematically, the plate approximation simply approximates to the actual smooth coefficients in the wave equation by a step function. While the latter may converge (both uniformly and absolutely!) to the former as the layering becomes finer and finer, so that the smooth equation is the limit of the layered one, it does not follow that the solution of the limiting equation will be the limit of the approximate equation! It is the old question of change of order in double limits. Actually one can construct counter-examples; for example, with an index linear in  $z$ , the wave equation can be solved by Bessel functions, and shows no back-reflection or other features of "fine layering".

The curvature and torsion of the L and the  $\Lambda$ -lines. As outlined in Section 6, under certain conditions a study of the differential geometry of an L- and  $\Lambda$ -line, mutually tangent at a point  $P_1: (x_1, y_1, z_1)$  -- "associated" with one another at that point, as we characterized them in that section -- will indicate their degree of coalescence with increasing circular frequency  $\omega$ . The condition generally given for such coalescence is (5.11), namely

$$(D.1) \quad \lim_{\omega \rightarrow \infty} \frac{1}{\omega^2} \frac{\nabla \cdot (\rho \nabla R)}{\rho R} = 0$$

Arthur D Little, Inc.

The travelling wave (A, B) which defines, by equations (5.5), the quantities R and S (within multiples of frequency  $\nu$ ) is supposed to have no branching singular locus within a distance D from  $P_0$ , which D does not diminish as  $\omega$  increases indefinitely--unlike the situation in the first of the examples given above.

We shall use a vector notation, in which  $\vec{P}$  of components (x,y,z) is a position vector of a point describing the L-line or the  $\Lambda$ -line, and in each case we use the arc-length s as the parametric variable. Then the tangent vector  $\vec{T}$  (of unit length, directed in the sense of increasing s) is given by the formula

$$(D.2) \quad \vec{T} = d\vec{P}/ds.$$

By definition of "curvature" K, and principal normal  $\vec{N}$ ,

$$(D.3) \quad K \vec{N} = d\vec{T}/ds = d^2 \vec{P}/ds^2.$$

Finally, the binormal  $\vec{B}$  being the common perpendicular to  $\vec{T}$  and  $\vec{N}$ , in a direction such that  $(\vec{T}, \vec{N}, \vec{B})$  is a vector triplet figure ("Frenet's trihedral") congruent with that formed by the unit vectors along the axes of coordinates, we have in terms of vector products

$$(D.4) \quad \vec{B} = \vec{T} \times \vec{N}.$$

An elementary account of these concepts and the needed formulas is given in such texts on analysis as W. F. Osgood,<sup>4</sup> Chapter VII. For a deeper treatment, such standard texts in differential geometry as L. P. Eisenhart and W. Blaschke may be consulted.<sup>4</sup> We shall assume these essentials here.

We wish to compare these quantities and vectors for the L and the  $\Lambda$ -lines at  $\vec{P}_1$ , and shall when necessary affix the corresponding subscript, L or  $\Lambda$ .

Consider first the case of rays, L, with  $\vec{P}$ ,  $\vec{T}$ ,  $\vec{N}$ ,  $\vec{B}$ , and K implicitly understood to have the subscript L. In Appendix A the ray differential equations (A.2) are written in terms of "time t" as the independent variable; to express them in the form needed here, with arc length s as the independent variable, we write  $ds = c dt = dt/n$ , whereupon these equations become, in vector notation,

$$(D.5) \quad \frac{d^2 \vec{P}}{ds} + \frac{dm}{ds} \frac{d\vec{P}}{ds} = \nabla m, \quad m = \log n = -\log c.$$

Note that  $dm/ds = \nabla m \cdot d\vec{P}/ds$ , and apply (D.2) and (D.3): then (D.5) assumes the form

$$(D.6) \quad K \vec{N} + (\nabla m \cdot \vec{T}) \vec{T} = \nabla m.$$

To find the curvature  $K = K_L$ , we take the scalar product of this equation with  $\vec{N}$ , applying the fact that  $\vec{N} \cdot \vec{N} = 1$  and that  $\vec{N} \cdot \vec{T} = 0$  (the latter, shown by differentiation of  $\vec{T} \cdot \vec{T} = 1$  and use of (D.3)). Then take the scalar product of this equation again with  $\nabla m$ . An obvious elimination in the two resulting equations gives:

$$(D.7) \quad K_L^2 = |\nabla m|^2 - (\nabla m \cdot \vec{T}_L)^2 = |\vec{T}_L \times \nabla m|^2,$$

the latter by an elementary vector formula (Lagrange's identity). Since the curvature is (by definition) essentially non-negative, (D.7) determines it completely in terms of the gradient of a physically determined function,  $c$ , and the tangent direction. The curvature thus being determined, (D.6) determines the principal normal  $N_L$  in terms of the same quantities. Finally, the binormal is found taking the vector product of (D.6) with  $\vec{T}$ , applying (D.4) and the fact that  $\vec{T} \times \vec{T} = 0$ . We obtain

$$(D.8) \quad K_L \vec{B}_L = \vec{T}_L \times \nabla m.$$

We turn now to the power flow lines,  $\Lambda$ . These, being normal to the loci  $S = \text{constant}$ , are directed along  $\nabla S$ , and have the direction vector

$$(D.9) \quad T_\Lambda = T = \frac{\nabla S}{|\nabla S|}$$

Using the arc  $s$  along  $\Lambda$  as parameter, we have by (D.2)

$$(D.10) \quad T_\Lambda = \frac{d\vec{P}}{ds} = \frac{\nabla S}{|\nabla S|} ;$$

hence the directional derivative of any function  $f(\vec{P}) = f(x, y, z)$  along  $\Lambda$  is given by

$$(D.11) \quad \frac{df}{ds} = \frac{d\vec{P}}{ds} \cdot \nabla f = \vec{T} \cdot \nabla f = \frac{\nabla S \cdot \nabla f}{|\nabla S|}$$

To apply (D.3) to  $\Lambda$ , we must find the value, in a useful form, of

$$(D.12) \quad \frac{d\vec{T}}{ds} = \frac{d}{ds} \frac{\nabla S}{|\nabla S|} = \nabla S \frac{d}{ds} \frac{1}{|\nabla S|} + \frac{1}{|\nabla S|} \frac{d}{ds} \nabla S.$$

For this purpose, we introduce the following quantity, analogous (as will be seen later) to  $m$ :

$$(D.13) \quad \mu = \log |\nabla S|.$$

For the first term on the right in (D.12) we have

$$\begin{aligned} \frac{d}{ds} \frac{1}{|\nabla S|} &= \frac{d}{ds} (|\nabla S|^2)^{-1/2} = -\frac{1}{2} |\nabla S|^{-3} \frac{d}{ds} |\nabla S|^2 \\ &= -\frac{1}{2} |\nabla S|^{-1} \frac{d}{ds} \log |\nabla S|^2 = -\frac{1}{|\nabla S|} \frac{d\mu}{ds} \end{aligned}$$

Here a free use has been made of (D.11). Thus we have

$$\nabla S \frac{d}{ds} \frac{1}{|\nabla S|} = -\vec{T} \frac{d\mu}{ds} = -\vec{T} (\vec{T} \cdot \nabla \mu).$$

To find the second term on the right in (D.12), we start with (D.13) and write

$$\nabla \mu = \frac{1}{2} \nabla \log |\nabla S|^2 = \frac{\nabla (|\nabla S|^2)}{2|\nabla S|^2}$$

The  $x$ -component of this gradient vector is found as follows:

$$\begin{aligned} \frac{1}{2|\nabla S|^2} \frac{\partial}{\partial x} (S_x^2 + S_y^2 + S_z^2) &= \frac{1}{|\nabla S|^2} (S_x S_{xx} + S_y S_{yx} + S_z S_{zx}) \\ &= \frac{1}{|\nabla S|} \left[ \frac{S_x}{|\nabla S|} S_{xx} + \frac{S_y}{|\nabla S|} S_{yx} + \frac{S_z}{|\nabla S|} S_{zx} \right] \end{aligned}$$

$$= \frac{1}{|\nabla S|} \frac{dS}{ds} \mathbf{x}$$

Similarly for the other two components. Thus we have

$$\nabla_{\mu} = \frac{1}{|\nabla S|} \frac{d}{ds} \nabla S,$$

which is the second term on the right in (D.12).

Introducing the values found in the last two paragraphs into (D.12), we obtain, on applying (D.3), the following equation (in which  $\Lambda$  subscripts are implied):

$$(D.14) \quad K \vec{N} + (\nabla_{\mu} \cdot \vec{T}) \vec{T} = \nabla_{\mu},$$

This is the exact analogue for  $\Lambda$  of (D.6) for  $L$ , the quantity  $\mu$  of (D.13) replacing  $m = \log n$  of the previous case. By precisely the same vector manipulations we obtain the following analogues of (D.7) and (D.8):

$$(D.15) \quad K_{\Lambda}^2 = |\nabla_{\mu}|^2 - (\nabla_{\mu} \cdot \vec{T}_{\Lambda})^2 = |\vec{T}_{\Lambda} \times \nabla_{\mu}|^2$$

$$(D.16) \quad K_{\Lambda} \vec{B}_{\Lambda} = \vec{T}_{\Lambda} \times \nabla_{\mu}$$

We pass now to the comparison of the  $L$  and the  $\Lambda$  curves, associated at  $P_1$ . From the definitions of  $m$  and  $\mu$ , we have

$$(D.17) \quad \begin{aligned} \mu - m &= \frac{1}{2} \log (c^2 |\nabla S|^2) \\ &= \frac{1}{2} \log \left[ 1 + \frac{c^2}{\omega^2} \frac{\nabla \cdot (\rho \nabla R)}{\rho R} \right] \end{aligned}$$

[illegible]

D-15

But even this is not sufficient: Even if the three unit vectors  $(\vec{T}, \vec{N}, \vec{B})$  forming the Frenet Trihedral approach coincidence as  $\omega \rightarrow \infty$ , and even though the rate of turning of  $\vec{N}_L$  and  $\vec{N}_\Lambda$  (the curvatures  $K_L$  and  $K_\Lambda$ ) approach coincidence, the torsion or rate of change in direction of the binormal per unit advance in arc-length, that is required:

$$(D.18) \quad \underline{\text{Torsion}} = \left| \frac{d\vec{B}}{ds} \right| = \left| \frac{d}{ds} \frac{\vec{T} \times \nabla \mu}{K} \right|.$$

From this and the earlier formulas, the approach to equality of the torsions is a second order condition: The partial derivatives of the second order of  $\mu - m$  must approach zero as  $\omega \rightarrow \infty$ .

While intuition based on Figure D.II would suggest that under these assumptions the curves would approach coalescence, this requires a further assumption to ensure the result by a rigorous mathematical proof: the rates at which the approach to zero of  $\mu - m$  and the first and second derivatives takes place must be uniform throughout the region  $W$  of xyz-space considered. This complete hypothesis: The uniform approach to zero of  $\mu - m$  and all its first and second order derivatives over  $W$ --which is equivalent to a third order condition on  $S$  or a fourth order one on  $R$ , as (D.17) shows--is our full hypothesis (5H) mentioned in Section 6 or, here, (DH).

Once this assumption is made, the coalescence of the  $L$  and  $\Lambda$  curves follows by a classical theorem in the differential geometry of space curves: If the curvature and torsion are given functions of arc-length along a curve, the curve is determined except for its position in space;



if a point  $P_1$  and a tangent direction  $\vec{T}_1$  are given, it is completely determined. This theorem is an immediate consequence of the uniqueness theorem in ordinary differential equations. A simple and elementary proof is given in Blaschke,<sup>4</sup> Vol. I, Chap. 1, Section 13. What we need is, as a matter of fact, the method of determining the curves: The differential equations (obtained from the Frenet formulas) have the curvature and torsion as coefficients. Thus two associated L and  $\Lambda$ -lines have the same initial conditions (mutually tangent at  $P_1$ ) and satisfy differential equations of the same form, and whose coefficients approach coincidence uniformly: Therefore, by the classical theory (Goursat-Hedrick,<sup>4</sup> Vol. II), their solutions do likewise.

It is possible to give our hypothesis (5H) a more elegant form in the case of analytic functions  $A$ ,  $B$ ,  $c$ --an assumption we are allowed to make in the usual cases, for the reasons set forth in Appendix C. The analyticity in the real domain  $W$  implies analyticity in a complex (6-dimensional) domain  $W+$  of the three coordinates regarded as complex variables. What happens then as  $\omega \rightarrow \infty$ ? It is possible that the domain  $W+$  shrinks up, although always containing its real part  $W$ ; it is also possible that the limit in (D.1) is not approached uniformly. On the other hand, if we assume that a complex domain  $W+$  exists which is fixed for all  $\omega$ , and that the convergence of  $\mu - m$  to zero (condition (D.1)) is uniform in  $W+$ --then, by the theory of analytic functions of complex variables, the derivatives of all orders of  $\mu - m$  will also approach zero uniformly within  $W+$ ; similarly, (5H) may be stated in the complex domain for the functions  $R_\omega$  and  $S_\omega$  themselves without mentioning their derivatives.

It is possible to visualize the condition in question: Usually,  $c$  changes very little with many meters of change in position. Hence, its derivatives are very small. With  $\mu$  having the relation of (D.H) to  $m$  means that all second and higher order derivatives of  $S$  are very small:  $S$  is almost linear in the coordinates, and, hence, the traveling wave is approximately a sine or cosine (or complex exponential) in a linear combination of the coordinates and the time.

APPENDIX E  
INTEGRAL INVARIANTS

A century and a half's experience in dealing with the mathematics of mechanical systems has established the basic importance of integral invariants. They had early applications to Hamiltonian systems of a finite number of degrees of freedom, and also to fluid dynamics. Their role shows itself to be crucial in statistical mechanisms, where it is now clear that whether one uses them consciously or not, they are unavoidable. The later sections of the present memorandum show the use of the integral invariants in the method of rays and ray statistics in hydroacoustic propagation. We may add, finally, that many branches of modern pure and applied mathematics (e.g., differential geometry, group theory, quantum mechanics) are increasingly incorporating the properties of integral invariants into their structure.

Examples. The most obvious example of an integral invariant is provided by the conservation of mass of a region  $R$  which is part of a deformable medium in motion, such as a flowing fluid. If the density and velocity components at the point  $(x, y, z)$  and time  $t$  are  $\rho$  and  $(u, v, w)$ , the flow corresponds with the equations

$$(E.1) \quad \frac{dx}{dt} = u, \quad \frac{dy}{dt} = v, \quad \frac{dz}{dt} = w,$$

and the mass of  $R$ --which we denote by  $R_t$  to show that it is moving with the flow--is

$$(E.2) \quad M(R_t) = \iiint_{R_t} \rho dx dy dz$$

It remains unchanged:  $M(R_t) = M(R_{t_0})$ , so that its derivative is zero. This is, of course, an alternative statement of the equation of continuity in either of its equivalent forms

$$(E.3) \quad \rho \cdot \frac{\partial(x, y, z)}{\partial(x_0, y_0, z_0)} = \rho_0$$

$$(E.4) \quad \frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} + \frac{\partial \rho w}{\partial z} = 0.$$

The constancy of the integral in (E.2)--for every choice of region--is described by saying that it is an integral invariant of the system of differential equations (E.1). Since there is no restriction on the region  $R_t$  (other than its "regularity": i.e., being a possible shape of a piece of matter), the integral invariant is said to be an absolute one. The equivalence of (E.2), (E.3), and (E.4) is shown by using the formulas of change of variables in definite integrals and of differentiation of the Jacobian determinants. Cf., e.g., W. F. Osgood, *Advanced Calculus*, Chapter XII and Chapter XIV, Sections 23-26. ( )

In addition to the absolute integral invariant (E.2), there is in the flow of an ideal fluid a relative integral invariant: in

order to remain unchanging with the flow it has to be taken about a closed manifold: the circulation  $ci(C_t)$  around the closed curve  $C_t$ , defined by the equation

$$(E.5) \quad ci(C_t) = \int_{C_t} u\delta x + v\delta y + w\delta z .$$

Here the symbol  $C_t$  denotes not only the curve moving with the fluid, but an arbitrarily assigned direction along it. The  $\delta$  is used for the differential along  $C_t$ --to avoid confusion with  $d$ , the differential along a particle's trajectory defined by (E.1).

The proof of the invariance of  $ci(C_t)$  depends on the Hamiltonian form which the equations of particle motion can be given in the case of an ideal fluid which is either incompressible or has its pressure a function of density and is acted on by body forces deriving from a force function  $V$  of  $(x, y, z, t)$ . For, on writing

$$(E.6) \quad H = H(x, y, z, t, u, v, w) = \frac{1}{2} (u^2 + v^2 + w^2) + V + \int \frac{dp}{\rho}$$

and using  $\partial$  to denote differentiation in the system of seven independent variables shown in  $H$  in (E.6), the equations of particle trajectories are

$$(E.7) \quad \frac{dx}{dt} = \frac{\partial H}{\partial u}, \quad \frac{dy}{dt} = \frac{\partial H}{\partial v}, \quad \frac{dz}{dt} = \frac{\partial H}{\partial w},$$

$$\frac{du}{dt} = -\frac{\partial H}{\partial x}, \quad \frac{dv}{dt} = -\frac{\partial H}{\partial y}, \quad \frac{dw}{dt} = -\frac{\partial H}{\partial z},$$

the first three coinciding with (E.1), the last three, with the standard hydrodynamical equations of conservation of momentum with the present type of fluid.

Note the contrast in this usage of  $\underline{\partial}$  with that in (E.4).

The first step in the application of (E.7) to the proof of the invariance of  $ci(C_t)$  is to express it as an integral about  $C_0$ , the position of the curve at some epoch taken as  $t = 0$ . We have

$$ci(C_t) = \int_{C_0} u\delta x + v\delta y + w\delta z;$$

but now the  $x, y, z$  must be thought of as functions of  $t$  and their "initial values"  $x_0, y_0, z_0$ , for  $t = 0$ --the initial point  $(x_0, y_0, z_0)$  tracing the curve  $C_0$  as  $(x, y, z)$  traces  $C_t$ , and  $\delta$  being along either  $C_0$  or  $C_t$ ; i.e., it is a differential with respect to some parameter along the curves (e.g., arc-length along  $C_0$ ). Since this parameter along the curve, and the time  $t$  along the particle trajectory, are independent variables, we have  $d\delta x = \delta dx$ , etc., this being the present form of the cross-derivative theorem.

The second step in the proof is the differentiation of the above expression with respect to  $t$ :

$$\begin{aligned} \frac{d}{dt} ci(C_t) &= \int_{C_0} \left( \frac{du}{dt} \delta x + \frac{dv}{dt} \delta y + \frac{dw}{dt} \delta z + u \frac{d}{dt}(\delta x) + v \frac{d}{dt}(\delta y) + w \frac{d}{dt}(\delta z) \right) \\ &= \int_{C_0} \left( \frac{du}{dt} \delta x + \frac{dv}{dt} \delta y + \frac{dw}{dt} \delta z + u \delta \left( \frac{dx}{dt} \right) + v \delta \left( \frac{dy}{dt} \right) + w \delta \left( \frac{dz}{dt} \right) \right) \end{aligned}$$

The third step is to apply (E.7); and afterwards, go back to  $C_t$  as the path of integration.  $dci(C_t)/dt$  becomes

$$\int_{C_t} \left( -\frac{\partial H}{\partial x} \delta x - \frac{\partial H}{\partial y} \delta y - \frac{\partial H}{\partial z} \delta z + u \delta \left( \frac{\partial H}{\partial u} \right) + v \delta \left( \frac{\partial H}{\partial v} \right) + w \delta \left( \frac{\partial H}{\partial w} \right) \right)$$

which is readily transformed into

$$(E.8) \quad \frac{d}{dt} ci(C_t) = \int_{C_t} \delta[(u^2 + v^2 + w^2)/2 - H].$$

Since  $C_t$  is closed and the bracket is single-valued, the integral reduces to zero, thus proving the invariance of the integral in (E.5).

The integral invariance of (E.5) is the basis for both Lagrange's theory of irrotational flow and velocity potential, and also for Helmholtz's theory of vortices.

In the important special case of steady motion, i.e., when the time does not enter explicitly into  $H$  in (E.6), absolute integral invariants exhibit a striking geometrical property, which we shall illustrate here in the case of the mass (E.2).

Since  $u, v, w$  no longer contain  $t$ , the equations (2.1) for the trajectories may be re-written as

$$(E.9) \quad \frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w},$$

which determine a fixed set of curves, the stream lines, one and only one through each point  $(x, y, z)$ : all fluid particles that ever pass through that point trace this same path, although at different epochs.

Let  $S$  be any piece of a smooth surface bounded by the curve  $C$ , and consider the class of stream-lines cutting it: they form the solid bundle shown in Figure E.1, where for simplicity of drawing,  $S$  is taken as small and not tangent to a stream-line.

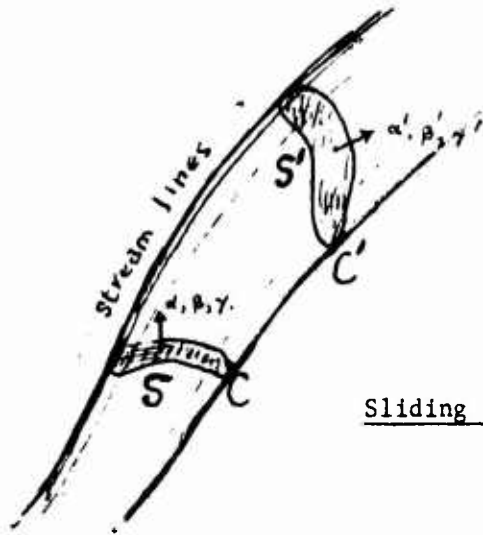


Figure E.1

Sliding Invariance of Mass Flux

The rate of passage of matter across  $S$  is evidently given by the surface integral over  $S$  of the normal component of the mass flux vector  $(\rho u, \rho v, \rho w)$ ; i.e., by

$$(E.10) \quad F(S) = \iint_S \rho(u \cos \alpha + v \cos \beta + w \cos \gamma) dS,$$

where  $\alpha, \beta, \gamma$ , are the direction angles of the normal to  $S$  in the sense shown.

Now consider the result of sliding all the points of  $S$  along the stream-line through them, by an arbitrary (but continuously and



smoothly varying) amount--in general different for the different points of S. In its new position, it is the surface S' bounded by C'; and the rate of passage of matter across it is

$$(E.11) \quad F(S') = \iint_{S'} \rho(u \cos \alpha' + v \cos \beta' + w \cos \gamma') dS.$$

Since with a steady flow there can be no change in the total amount of matter in a given region, and since no matter crosses the tube formed by the stream-line through C and C', we must have  $F(S') = F(S)$ : The integral in (E.10) is invariant under the operation of sliding along the stream-lines.

We shall see later how (E.11) corresponds to a relative sliding integral invariant about a closed curve generalizing the circulation about the boundary C' of S'--but now without the restriction of simultaneously (a fixed value of t in the calculation of the integrand over C').

The general theory of integral invariance in the time-unchanging sense was founded by H. Poincaré,<sup>5</sup> who was interested in Hamiltonian dynamical systems of n degrees of freedom, in whose 2n-dimensional phase-space of coordinates and momenta (q, p) the equations of motion

$$(E.12) \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$$

establish a "flow". By applying the  $2n$ -dimensional extension of the equation of continuity (E.4), with  $\rho = 1$  and (E.12) replacing (E.1), it is seen that this flow is incompressible: the  $2n$ -fold integral  $\int \int^n dq_1 dp_1 \dots dq_n dp_n$  over any flowing region  $R_t$  is constant. This is Liouville's theorem, an essential building block of statistical mechanics, as noted earlier. It is also shown that the "circulation" in phase space, i.e.,

$$(E.13) \quad \int_{C_t} \sum_1 p_1 \delta q_1$$

is also a relative integral invariant--by essentially the proof outlined in the hydrodynamic case given above. From (E.13) a sequence of absolute integral invariants of increasing dimensionality are derived. All this is valid for a general Hamiltonian  $H$  containing the time explicitly, and compares integrals over manifolds at one and the same epoch  $t$  with what they become at an arbitrary but fixed later epoch  $t'$ : the invariance may be described as kinematical.

E. Cartan<sup>5</sup> took up the whole subject from a different point of view in which the invariance of the integrals is more geometrical (but in a higher dimensional space): the points of the manifolds of integration are all slid along the integral curves, each an arbitrary amount, as described above in connection with Figure E.I. The invariance property may be described as sliding invariance. Cartan showed that--under a possible change of dimensionality--the kinematical and the sliding invariance are coextensive: kinematical invariance in

N-dimensional "space" becomes sliding invariance in  $(N + 1)$ -dimensional "space-time" of a related integral, easy to construct from the original one; and reciprocally, every sliding invariant leads to a (truncated) kinematic invariant in space of one dimension lower, the removed dimension being interpreted as "time".

Not only has the sliding invariant point of view greatly simplified the mathematical treatment, but it is the form directly applicable to acoustic rays. Simple as such applications may be at the conceptual level--provided one has the habit of thinking in terms of a "space" of  $n$  coordinates and of smooth manifolds embedded in them--the formulas, algebraic calculations and proofs of the sort found in the earlier literature are excessively complicated. As shown by E. Cartan and his followers, this is because the traditional notation (based on partial derivatives and determinants) is a misfit. A notation based on differentials and their skew-products not only provides simpler formulas, but leads to powerful algorithms.

Manifolds and oriented manifolds. In the physical applications of present concern, integrals over  $k$ -dimensional manifolds in  $n$ -dimensional space ( $n \geq k$ ) are of constant occurrence; and so are applications to them of the theorems of Gauss and Stokes and their generalizations. When integrating expressions having physical meaning over manifolds, such as  $S_k$ , two different cases are encountered:  $S_k$  may be a geometrical figure ( $k$ -dimensional spread of points); or it may be a different object entirely: a figure provided with an orientation. The two kinds of

integrands reflect the two situations. We illustrate this contrast in the familiar cases of  $S_1$  (a curve) and  $S_2$  (a surface) in ordinary xyz-space. Questions having definite answers when these figures are not oriented are, for instance, what is the length of  $S_1$ ? The area of  $S_2$ , or its mass (if it is a material of surface density  $\sigma$  given at each point)? The first is answered by integrating  $\sqrt{dx^2 + dy^2 + dz^2}$  over  $S_1$ ; the second and third by integrating the square root of a sum of squares of Jacobian determinants, or such a square root multiplied by  $\sigma$  to give the mass.

Quite different is the case in which we have a vector field of components  $(X, Y, Z)$  throughout our manifold, and (if it is a field of force) we ask for the value of the work done by this force on a particle moving along the curve  $S_1$ ; or else (if it is a field of velocities) we ask for the value of the flux through the surface  $S_2$ : neither question has a definite answer, unless we regard  $S_1$  and  $S_2$  as being oriented figures:  $S_1$  must mean a curve as well as a direction along it (a "path"); and  $S_2$  must be a surface with an orientation, so that we can say in which direction the flow takes place. In both cases the numerical answer will be algebraic: positive, negative, or zero. The requirement of orientability of surfaces and higher-dimensional manifolds imposes a restriction: the area, or mass of a material Mobius strip is meaningful, and so is the volume of fluid that encounters it per unit time; but not the volume that traverses it--in the ordinary usage of this term.

The orientation of manifolds is a less obvious notion in spaces of higher dimension than two or three; e.g., in the space of the variables  $(x, y, z, p_x, p_y, p_z)$  used so frequently in the text, along with the manifold which is the locus of the equation  $H = 0$ . In such cases, the concept of the indicatrix is used. It is easily explained in the spaces that can be visualized--and has the advantage of being extendable to a  $k$ -dimensional manifold in  $n$ -dimensional space.

The indicatrix on a curve  $S_1$  resembles an arrow-head that marks the direction on the curve: it is a pair of distinctly labeled points, e.g., marked (1, 2), with the direction from the first to the second. An essential part of the idea is that the pair may be slid along  $S_1$ , but must always remain "close together"--a restriction that can be stated precisely in terms of neighborhoods of points on  $S_1$ . This "smallness" of the indicatrix is needed to exclude the possibility of reversing it in the case that  $S_1$  is closed (e.g., a circle), when otherwise (1, 2) could be brought into coincidence with (2, 1) and would not determine a unique orientation of  $S_1$ .

In the case of a 2-dimensional manifold  $S_2$  (a surface) the usual method of orientation by defining the positive sense of the normals is applicable only when  $S_2$  lies in 3-dimensional space. It fails in higher dimensions: even in 4-dimensional space there is a continuous infinitude of normals to  $S_2$  at each one of its points (cf. the normals to a curve in 3-space). On the other hand, a "very small" closed oriented curve in the surface  $S_2$ , which is allowed to slide

about in it while remaining "arbitrarily small", reveals the two following possibilities: if, as in a sphere, the indicatrix can never reverse its sense, we say that the surface is orientable and define its orientation in any particular case by the orientation of the indicatrix. If on the contrary, as in the case of the Mobius strip, this reversal can take place, we say that the surface is non-orientable. These examples illustrate two facts: first, that orientability as defined by the indicatrix is an internal property of the manifold, and does not depend on the space in which it may lie; second, that the definition can be extended by induction from lower to higher dimensional manifolds. Thus an indicatrix  $I_k$  in  $S_k$  is defined firstly as a small "sphere-like"  $(k-1)$ -dimensional manifold (one which encloses a part of  $S_k$ , and which, with its interior, can be put into a one-to-one continuous correspondence with the set  $K_k$  of points in the space of  $(x_1, \dots, x_k)$  for which  $x_1^2 + \dots + x_k^2 \leq 1$ )--and secondly as this sphere-like manifold provided with an orientation (it is easily shown to be orientable). Using this  $I_k$  we define orientability, or its reverse, for  $S_k$ , and a particular orientation in the former case: this is our inductive process.

The rigorous background of these notions are found in any modern work on elementary topology.

The general differential notation. Returning to the first example given for the oriented curve (path)  $S_1$ , the work done by the field  $(X, Y, Z)$  is the line integral of the tangential component of the force in the direction in which  $S_1$  is oriented. In terms of the

direction cosines in this direction, the work is given by any one of the three expressions (ds being the arc differential in the direction of  $S_1$ )

$$\begin{aligned} & \int_{S_1} (X \cos \alpha + Y \cos \beta + Z \cos \gamma) ds \\ &= \int_{S_1} \left( X \frac{dx}{ds} + Y \frac{dy}{ds} + Z \frac{dz}{ds} \right) ds \\ &= \int_{S_1} (X dx + Y dy + Z dz) \end{aligned}$$

The second is derived from the first by the direction cosine equations,  $\cos \alpha = dx/ds$ , etc; and the third, by cancellation of ds--which is permitted by the properties of differentials. It is this third expression which is particularly interesting to us: it has the advantage of being free from any particular parameter in the representation of the oriented curve  $S_1$ . Thus if any other parameter,  $t$ , were used, chosen to increase in the direction of the orientation of  $S_1$ , which is given by such equations as

$$S_1: \quad x = f(t), \quad y = g(t), \quad z = h(t),$$

it would only be necessary to integrate  $Xf'(t) + Yg'(t) + Zh'(t)$  respect to  $t$  between appropriate limits (after replacing  $(x, y, z)$  in the coefficients by their expressions in  $t$ ) to obtain the correct value of the work. This independence of parametrization--which is not, after all, intrinsic in the nature of the path  $S_1$  regarded as a physical entity--is the advantage of the differential notation for the integrand. It may be compared with the freedom from particular coordinate systems achieved by the vector notation. It is an advantage that we wish to

exploit in higher dimensional cases,  $S_k$ .

In the example of the flow through the oriented surface  $S_2$ ,  $(X, Y, Z)$  being the velocity components, let us first define the positive normal as the one about which the indicatrix cycle in  $S_2$  corresponds to positive screw rotation. If  $\cos \alpha, \cos \beta, \cos \gamma$  are the direction cosines of this positive normal, the flux through  $S_2$  (volume traversing it in this positive direction per unit time) is given by

$$F(S_2) = \iint_{S_2} (X \cos \alpha + Y \cos \beta + Z \cos \gamma) dS$$

Here  $dS$  is the surface element of area, regarded as essentially positive.

Suppose that  $S_2$  is given parametrically in terms of the parameters  $u, v$ :

$$x = f(u, v), \quad y = y(u, v), \quad z = h(u, v);$$

and suppose, further, that rotation in the positive sense in the  $uv$ -plane (from  $+u$  to  $+v$ ) corresponds with circulation in the sense of the directrix of the corresponding points in  $S_2$ . Then by elementary differential geometry (cf. Osgood, Chapter VI), and as explained in Figure A.I and in the derivation of equation (A.6) in Appendix A (with  $\theta = u, \phi = v$ ), we have for the element of area  $dS = D du dv$ , where  $du dv$  is the corresponding element of area in the  $uv$ -plane (essentially positive) and  $D$  is the (essentially positive) length of the vector product  $(\partial \vec{r} / \partial u) \times (\partial \vec{r} / \partial v)$ . Furthermore,  $\cos \alpha, \cos \beta, \cos \gamma$



are the components of this product, divided by D. Now the components of this vector product are the three determinants (with the usual cyclical convention of sign) in the matrix (subscripts denoting partial differentiation)

$$\begin{vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix}$$

Therefore

$$\cos \alpha = \frac{\partial(y, z)}{\partial(u, v)} \cdot \frac{1}{D}, \quad \cos \beta = \frac{\partial(z, x)}{\partial(u, v)} \cdot \frac{1}{D}, \quad \cos \gamma = \frac{\partial(x, y)}{\partial(u, v)} \cdot \frac{1}{D}$$

Introducing these expressions into the earlier formula for the flux, we have, after cancelling D,

$$F(S_2) = \iint_{S_2(u, v)} \left[ x \frac{\partial(y, z)}{\partial(u, v)} + y \frac{\partial(z, x)}{\partial(u, v)} + z \frac{\partial(x, y)}{\partial(u, v)} \right] du dv$$

Once  $(x, y, z)$  are replaced by their expressions in  $(u, v)$ , and region  $S_2(u, v)$  in the  $uv$ -plane which is mapped onto  $S_2$  is found, the evaluation of  $F(S_2)$  reduces to double integration of a function of  $(u, v)$  over  $S_2(u, v)$ .

In spite of appearances, this formula has a form independent of the parametrization: in a different parametrization,  $(u, v)$  would be single valued differentiable functions of  $(u', v')$ --and vice versa. By the rules of change of variables in partial derivatives, and of linear transformations of determinants, the result would have exactly

the same form in the new parameters ( $u'$ ,  $v'$ ) as in the old. In view of this fact, mathematicians have sought to define a notation with a form independent of parametrization; with results to be given here.

Consider, for example, the coefficient of  $Z$ ; if  $du dv$  is a product of  $du$  and  $dv$  (as is the case with a net of rectangles on the  $uv$ -plane) we could write the coefficient in the form

$$\frac{\partial(x, y)}{\partial(u, v)} du dv = \begin{vmatrix} x_u du & y_u du \\ x_v dv & y_v dv \end{vmatrix} = \begin{vmatrix} dx & dy \\ d'x & d'y \end{vmatrix}$$

where  $d$  is the differential along a  $u$ -curve ( $v$  constant) and  $d'$ , along a  $v$ -curve ( $u$  constant). This has suggested the abbreviation  $[dx dy]$  for the determinant; it is a sort of product, but since  $[dy dx] = -[dx dy]$  it behaves skew-symmetrically. Notations have varied in the recent period, but are now fixed to read  $dx \wedge dy$ , and called the exterior product.

With this notation, we may write

$$F(S_2) = \int \int_{S_2} (X dy \wedge dz + Y dz \wedge dx + Z dx \wedge dy)$$

which, by definition, calls for the following set of operations, starting with the given  $(X, Y, Z)$  and oriented  $S_2$ :

1. Express  $S_2$  parametrically in terms of any parameters  $(u, v)$  mapping the positive indicatrix in  $(u, v)$  into the indicatrix of orientation of  $S_2$ .

2. Replace  $(x, y, z)$  in the coefficients by their parametric expressions in  $(u, v)$ .
3. Replace each exterior product,  $dy \wedge dz$ , etc., by the corresponding Jacobian determinant  $\partial(y, z)/\partial(u, v)$ , times the essentially positive element of area  $du dv$  of the  $uv$ -plane.
4. Integrate the result over the region  $S_2(u, v)$  of the  $uv$ -plane which corresponds to  $S_2$ , with respect to the variables of double integration,  $(u, v)$ .

This convention is general. If  $S_k$  is an oriented  $k$ -dimensional manifold in  $n$ -space, the integral over  $S_k$  of a sum of terms such as  $A dx_1 \wedge \dots \wedge dx_k$  is to be interpreted in terms of functional determinants: each exterior product  $dx_1 \wedge \dots \wedge dx_k$  is to be replaced by the functional determinant  $\partial(x_1, \dots, x_k)/\partial(u_1, \dots, u_k)$ , where  $(u_1, \dots, u_k)$  are parameters figuring in the (local) representation of  $S_k$ . The result is then integrated with respect to these in the parameter space. In the case of a sphere (and many other surfaces) a single parametrization is impossible: it has to be divided into pieces, each of which can be given in terms of an individual parameter system. In the map from any of these (local) parameter systems to the corresponding piece of  $S_k$ , orientation is preserved.

The exterior products can be manipulated algebraically: multiplied by a number (or function), added, and multiplied together, provided that any change of order of adjacent factors is understood to

reverse the sign of the product. Thus their properties parallel those of determinants. In particular, if two differentials in such a product are the same, the product must vanish.

There is a further operation, belonging to the calculus rather than to algebra, that is important in the physical applications, that of differentiation, defined as follows:

Let  $\Omega$  have the form of the integrand on an oriented manifold  $S_k$ : a sum of the type

$$\Omega = \sum A \, dx_1 \wedge \dots \wedge dx_k$$

We define the exterior differential  $d\Omega$  as follows

$$d\Omega = \sum dA \wedge dx_1 \wedge \dots \wedge dx_k$$

Working out the differential  $dA = \sum_{i=1}^m (\partial A / \partial x_i) dx_i$  we have

$$d\Omega = \sum \sum_{i=1}^m \frac{\partial A}{\partial x_i} \, dx_i \wedge dx_1 \wedge \dots \wedge dx_k$$

and removing all terms with a  $dx_i$  coinciding with one of  $dx_1, \dots, dx_k$ , we get an expression similar to  $\Omega$  but of one degree higher, and whose coefficients are first derivatives of those of  $\Omega$ . It is easy to see that the formalism of differentiation of the calculus extends to this exterior differential.

To interpret  $d\Omega$  in familiar cases, return to the first example (work) in which the integrand had the form

$$\Omega = X dx + Y dy + Z dz.$$

According to the above definition we have

$$d\Omega = dX \wedge dx + dY \wedge dy + dZ \wedge dz$$

which becomes (subscripts denoting partial derivatives)

$$d\Omega = (Z_y - Y_z) dy \wedge dz + (X_z - Z_x) dz \wedge dx + (Y_x - X_y) dx \wedge dy.$$

Evidently the coefficients are the curl of the vector  $(X, Y, Z)$ , while, as before, the exterior products are the direction cosines of the normals, times the element of area, on any orientable surface of which  $S_1$  is the complete contour. This is the basis of Stokes' theorem.

As a second illustration, consider the flux through  $S_2$  in the velocity field  $(X, Y, Z)$ . On writing the integrand in  $F(S_2)$  in the form

$$\Omega = X dy \wedge dz + Y dz \wedge dx + Z dx \wedge dy$$

we have

$$\begin{aligned} d\Omega &= dX \wedge dy \wedge dz + dY \wedge dz \wedge dx + dZ \wedge dx \wedge dy \\ &= (X_x + Y_y + Z_z) dx \wedge dy \wedge dz \end{aligned}$$

which is the basis of the divergence theorem of Gauss.

All the straight-forward generalizations of these theorems are expressible in the notation of exterior differentials and differentiation. We note only two facts, the first being easy to prove, namely, if  $\Omega$  is itself a differential:  $\Omega = d\omega$  then  $d\Omega = 0$ . The second fact is much harder to prove; it is a converse to the former: if  $d\Omega=0$ , then  $\Omega = d\omega$ , but the coefficients in  $\omega$  may be multiple-valued--the theorem as a simple converse must be regarded as a local theorem.

Some general theorems. We wish first to replace the relative integral invariant (E.13) of the Hamiltonian system (E.12) by a sliding invariant, namely

$$(E.14) \quad \int_K \left( \sum_1 p_1 \delta q_1 - H \delta t \right),$$

where  $K$  is a closed curve in the  $(2n + 1)$ -dimensional "phase-space-time" of the  $n$  variables of position ( $q$ ) of a system of  $n$  degrees of freedom, the  $n$  corresponding momenta ( $p$ ), and the time  $t$ . Now we must think of  $K$  as given by the parametric equations in the parameter  $u$ :

$$K: q_1 = q_1(u), \quad p_1 = p_1(u), \quad t = t(u), \quad (0 \leq u \leq h)$$

where, since  $K$  is closed, the values of these  $2n + 1$  functions at  $u = h$  are the same, respectively, as at  $u = 0$ . The differential  $\delta$  is along  $K$ ; i.e., it implies that  $u$  is the independent variable.

To express the operation of sliding along the integral curves of the Hamiltonian equations of motion (E.12), we require a second parameter,  $v$ ; then the above equations are replaced by the following:

$$(E.15) \quad K_v: \quad q_1 = q_1(u, v), \quad p_1 = p_1(u, v), \quad t = t(u, v).$$

The differential  $\underline{d}$  will be used to indicate the sliding of the points of  $K = K_v$  with increase of  $v$ . Since the paths along which the sliding occurs must be the integral curves of (E.12), the direction tangent to them in  $(2n + 1)$ -space (of components  $dq_1, dp_1, dt$  in (E.15),  $u$  being held fast) must coincide with the direction satisfying (E.12)--which equations may therefore be written in the form

$$(E.16) \quad \frac{dq_1}{Hp_1} = \dots = \frac{dq_n}{Hp_n} = \frac{dp_1}{-Hq_1} = \dots = \frac{dp_n}{-Hq_n} = \frac{dt}{1} = dv.$$

As in the earlier case, the interpretation of the two systems of differentials in terms of the variables in (E.15) shows that their order of application is interchangeable: for any function  $f$  of these  $2n + 1$  dependent variables,  $d\delta f = \delta df$ .

With these preliminaries, the proof is quite similar to that of the invariance of circulation given earlier. The integral (E.14) becomes--when  $K = K_v$ , the path defined by (E.15)--a function of  $v$ , which function we must show to be a constant; i.e., to have a zero  $\underline{d}$  differential. Thinking of the variables in the integrand of (E.14) as standing for their values given in (E.15), the integrand becomes a function of  $(u, v)$ , which is integrated with respect to  $u$  between the limits  $u = 0$  and  $u = h$ ,  $v$  being regarded as a constant. The result is a function of  $v$ , and our task is to show that its  $\underline{d}$  differential ( $v$  varying) is zero. By Leibnitz' rule, we may take the  $\underline{d}$  under the

integral sign, so that what we are concerned with is the d differential of the integrand in (E.14).

The calculations are simplified by use of the notation

$$(E.17) \quad \omega(\delta) = \sum_i p_i \delta q_i - H \delta t, \quad \omega(d) = \sum_i p_i dq_i - H dt.$$

We wish to prove the equation

$$(E.18) \quad d\omega(\delta) - \delta\omega(d) = 0,$$

on the basis of (E.15) and (E.16). Using the latter, we have

$$\begin{aligned} \delta H dt &= dt \sum_i (H p_i \delta p_i + H q_i \delta q_i) + H_t \delta t dt \\ &= \sum_i (dq_i \delta p_i - dp_i \delta q_i) + H_t \delta t dt \\ dH \delta t &= 0 + H_t dt \delta t \end{aligned}$$

The truth of (E.18) follows from the application of the last two equations, together with the interchangability of order:  $d\delta = \delta d$ .

Returning to (E.14), and using the results just obtained, we have

$$d \int_{u=0}^{u=h} \omega(\delta) = \int_{K_v} d\omega(\delta) = \int_{K_v} \delta\omega(d) = \omega(d) \Big|_{u=0}^{u=h} = 0$$

since  $K$  is closed and the integrand single valued. This completes the proof.

Note that the left-hand member of (E.18) is the exterior



differential of  $W$ ; it can be written

$$\Omega = dp_1 \wedge dq_1 - dH \wedge dt$$

We have shown that if in (E.15)  $v$  varies along the path,  $\Omega = 0$ . If on the contrary (E.15) are regarded as defining a piece of surface  $S_2$ --not in general either closed or tangent to any integral curve of (E.12)--then  $\Omega$  is not necessarily zero. Our next theorem is that  $\Omega$  is the integrand of an absolute sliding integral invariant of (E.12).

This theorem would follow immediately from the invariance of (E.14) by the application of the  $(2n + 1)$ -dimensional generalization of Stokes' theorem to the piece of surface defined by (E.15) in their present interpretation. But since we do not wish to assume this generalization here, we shall outline a direct analytic proof, quite similar to the one given above. We must introduce into the functions of  $(u, v)$  in (E.15) a third parameter,  $w$ , which varies along the integral curves, and hence, when  $dv$  is replaced by  $dw$ , (E.16) are satisfied. We are to show that if  $\Omega$  is integrated with respect to  $(u, v)$  over  $S_2$  the  $w$ -differential is zero. Again Leipnitz' rule applies, and there is no real trouble in justifying the following formal steps:

$$d \int \int_{S_2} \Omega = \int \int_{S_2} d\Omega = 0$$

The fact that  $d\Omega = 0$  follows from the form of  $\Omega$  and differential manipulations involving (E.16).

Our last class of results are based on the following theorem  
of E. Cartan:

The integral  $\int_{S_k} \Omega$  will be a sliding invariant of a system of differential equations (such as (E.16), or, more generally (9.1) of Section 9) if and only if  $\Omega$  can be expressed in terms of the first integrals of this system and their differentials.

This is a local theorem (i.e., valid only in a sufficiently small neighborhood of a point of regular behavior of the differential equations). The formal proof involves the elementary properties of the  $N-1$  first integrals of a differential system of order  $N$ . [In (E.16)  $N = 2n + 1$ ; in (9.1),  $N = n$ ]. Described intuitively, we "deform" the neighborhood in question so as to "straighten out" its integral curves in this neighborhood. Analytically, this amounts to introducing the  $N-1$  first integrals as independent variables (denote them by  $y_1, \dots, y_{N-1}$ ) and any  $N$ 'th function, independent of the latter, as the last variable  $y_N$ .

In the space of  $(y_1, \dots, y_N)$  the integral curves are straight lines, one and only one through each point, and all parallel to the  $y_N$  - axis. The integrand  $\Omega$  becomes an expression in terms of  $(y_1, \dots, y_N)$  and their differentials. Since its integral over any  $S_k$  in the space of these variables has a value that is unchanged by any sliding along lines parallel to the  $y_N$  - axis, the integrand  $\Omega$  must contain neither this variable nor its differential. It is expressible

solely in terms of  $(y_N, \dots, y_{N-1})$  and their differentials. These being the first integrals of the differential equations, Cartan's theorem is proved.

From Cartan's theorem it follows that from any two sliding invariants, a third can be obtained by exterior multiplication. Thus if  $\Omega$  and  $\Omega'$  are two integrands in our standard (exterior product) form,  $\Omega \wedge \Omega'$  is an integrand of the same form. For if each factor is expressible in terms of the first integrals and their differentials, the same will be true of this product. Applying this to the case in which

$$\Omega = \Omega' = d\left(\sum_1 p_1 dq_1 - H dt\right) = \sum_1 dp_1 \wedge dq_1 - dH \wedge dt ,$$

we obtain a whole sequence of sliding absolute integral invariants of even orders: 2, 4, 6, ..., 2n. The last one has an integrand reducing to that in Lionville's theorem (volume preservation in phase space) when it is specialized to the 2n-dimensional phase space with  $t = \text{constant}$ .

Another consequence of Cartan's theorem is that if any  $\Omega$  is the integrand of a sliding invariant, so is its exterior differential  $d\Omega$ . This, naturally, gives nothing new, when, as in our 2-dimensional Hamiltonian invariant,  $d\Omega = 0$  identically. On the other hand, a rather simple extension of Cartan's theorem applies to relative sliding invariants: It is a straightforward matter to show that their integrands can be expressed in terms of the first integrals and their differentials

to which is added a perfect differential,  $dG$ , of a function  $G$  of all the coordinates: When these are  $y_1, \dots, y_N$ , the (non-integral)  $y_N$  will be involved. Therefore, we would have, for our  $\omega(\delta)$  defined before:

$$W(\delta) = \sum_{i=1}^{N-1} f_i(y_1, \dots, y_{N-1}) \delta y_i + \delta G(y_1, \dots, y_N)$$

Taking the exterior differential, the  $\delta G$  term yields zero, so that  $dW$

$$dW = \sum_{i=1}^{N-1} df_i \wedge dy_i$$

and is, therefore, the integrand of an absolute sliding integral invariant. This is an alternative proof of the second theorem, the proof of which was outlined above.

A final application of Cartan's theorem shows how, from a given sliding integral invariant of highest order, and given a first integral of the differential system, we can obtain such an invariant, of order one unit lower, valid on each manifold defined by setting this first integral equal to a constant.

Let  $\omega$  be such an absolute integral invariant of order  $N-1$ , and let  $Y(x_1, \dots, x_N)$  be such a first integral. We may take it in our complete set of  $N-1$  first integrals, writing  $y_1 = Y$ , and  $y_2, \dots, y_{N-1}$  being the remaining integrals. Then in these variables,

$$\omega = L(y_1, \dots, y_{N-1}) dy_1 \wedge dy_2 \wedge \dots \wedge dy_{N-1}$$

So far all that we know about  $L$  is that it exists, and that, setting

$y_1 = Y = y_1^0$ , any constant, the expression  $L(y_1^0, y_2, \dots, y_{N-1}) dy_2 \wedge \dots \wedge dy_{N-1}$

has the invariant form required. To find its expression, we must start with what we are given explicitly, namely  $Y$  as a function of  $(x_1, \dots, x_N)$ , and  $\omega$  in terms of these coordinates and their differentials, i.e., as

$$A_1 dx_2 \wedge \dots \wedge dx_N + A_2 dx_1 \wedge dx_3 \wedge \dots \wedge dx_N + A_N dx_1 \wedge \dots \wedge dx_{N-1}$$

Let us introduce  $Y$  as a new coordinate, expressing one of the  $x$  coordinates in terms of it. Suppose, e.g., that  $Y_N = \partial Y / \partial x_N \neq 0$ : We can solve the equation  $Y(x_1, \dots, x_N) = y_{N-1}$  for  $x_N$  in terms of  $y_{N-1}$  and  $x_1, \dots, x_{N-1}$ . Replacing  $x_N$  by this expression and  $dx_N$  by

$$dx_N = dy_{N-1} / Y_N - (Y_1 dx_1 + \dots + Y_{N-1} dx_{N-1}) / Y_N$$

in the given expression for  $\omega$ , it becomes

$$\frac{1}{Y_N} (A_1 dx_2 \wedge \dots \wedge dx_{N-1} + A_2 dx_1 \wedge dx_3 \wedge \dots \wedge dx_{N-1} + \dots + A_{N-1} dx_1 \wedge \dots \wedge dx_{N-2}) \\ \wedge dy_{N-1} + \omega'$$

where  $\omega'$  does not contain  $dy_{N-1}$ . Equating this to the earlier expression,  $L dy_1 \dots dy_{N-1}$  (with the above choice of  $y_{N-1}$ ), and taking the exterior product by  $dy_{N-1}$  of the resulting equation, we get the identity  $\omega' \wedge dy_{N-1} = 0$ , and since  $\omega'$  does not have  $dy_{N-1}$  as a factor, this identity in the  $N$  differentials implies that  $\omega' = 0$ .

The result of this is that the difference between the two expressions

$$L(y_1, \dots, y_{N-1}) dy_1 \wedge \dots \wedge dy_{N-2}$$

and

$$\Omega = (A_1 dx_2 \wedge \dots \wedge dx_{N-1} + \dots + A_{N-2} dx_1 \wedge \dots \wedge dx_{N-2}) / Y_N$$

when multiplied exteriorly by  $dy_{N-1}$ , is identically zero. This means that the difference must either vanish identically, or else contain  $dy_{N-1}$  as an (exterior) factor, as a rather simple manipulation of such "exterior polynomials" shows. In either case we see that  $\Omega$ , on each manifold  $Y = \text{constant}$ , reduces to our expression containing only the first integrals  $y_1, \dots, y_{N-2}$  and their differentials, so that its integral is an absolute sliding invariant.

This process is the general and rigorous parallel of the intuitive discussion at the end of Section 9, based on Figure 9.I in the case  $N = 3$ . It can be applied repeatedly as further first integrals are given. In the case of Lionville's theorem obtained as described above, there is an evident application when  $H$  does not contain the time explicitly. Then  $H$  is a first integral, so that the coefficient of  $dH$  in the  $2n$ -dimensional invariant is the exterior product of  $dt$  times a  $(2n-2)$ -dimensional expression, which is readily shown to be the integrand of a new integral invariant--of fundamental importance in hydro-acoustics. In this particular case, since our manifold is characterized by the equation  $H = 0$ , there is a simpler direct way of deriving it.

When we confine ourselves to the 5-dimensional locus  $H = 0$  in the 7-dimensional phase-space-time the integrand of the absolute integral invariant

$$\Omega = dp_x \wedge dx + dp_y \wedge dy + dp_z \wedge dz - dH \wedge dt = dp_x \wedge dx + dp_y \wedge dy + dp_z \wedge dz$$

which has been used in Sections 13, 14, and 15 in the case of propagation in a vertical plane (aximuthal symmetry).

Taking the exterior product  $\Omega^2 = \Omega \wedge \Omega$ ,

$$\Omega^2/2 = dx \wedge dy \wedge dp_x \wedge dp_y + dx \wedge dz \wedge dp_x \wedge dp_z + dy \wedge dz \wedge dp_y \wedge dp_z$$

which served as the basis of the treatment of the general case in Sections 11 and 12. Its geometrical representation is shown in Figure 11.I, and its integral  $I_4$  is evaluated in (11.6), etc.

## APPENDIX F

### AN INTEGRAL SOLUTION OF LINEAR DIFFERENTIAL SYSTEMS

The system of six non-homogeneous linear differential equations (12.56) of Section 12 are of the type (using  $x$  for the independent variable and  $y_1, \dots, y_n$  for the unknown functions):

$$(F.1) \quad dy_i/dx = y_1 q_{1i} + \dots + y_n q_{ni} + r_i \quad (i = 1, \dots, n),$$

where the coefficients  $q_{ji}$  and the non-homogeneous terms  $r_i$  are known functions of  $x$ . In the case of Section 12,  $n = 6$  and the line-up of letters is

$$\begin{array}{ccccccc} x & y & z & p_x & p_y & p_z & \tau \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & x \end{array}$$

By "one solution" is meant such a horizontal row of  $n$  functions.

It is known that there exists a "fundamental system" of solutions of the homogeneous equations to which (F.1) would reduce if every  $r_i$  were replaced by zero. It is a set of  $n$  solutions, unconnected by any linear relation with constant coefficients, and in terms of which every other solution of the homogeneous equations can be expressed as a homogeneous linear combination with constant coefficients.

Let the  $j$ 'th solution in this fundamental system be denoted by  $(z_{j1}, \dots, z_{jn})$ , so that (for each  $j$ ) the following homogeneous equations are satisfied

$$(F.2) \quad dz_{j1}/dx = z_{j1} q_{11} + \dots + z_{jn} q_{n1}, \quad (i = 1, \dots, n).$$



These relations as well as the later manipulations are most easily expressed in matrix form. First we introduce the one-row matrices (n-vectors)

$$Y = \begin{vmatrix} y_1 & \dots & y_n \end{vmatrix}$$

$$R = \begin{vmatrix} r_1 & \dots & r_n \end{vmatrix}$$

The coefficients in (F.1) are written as the square matrix

$$Q = \begin{vmatrix} q_{11} & \dots & q_{1n} \\ \vdots & & \vdots \\ q_{n1} & \dots & q_{nn} \end{vmatrix}$$

The n solutions in our fundamental system form n rows of n variables; we write them as the square matrix

$$Z = \begin{vmatrix} z_{11} & \dots & z_{1n} \\ \vdots & & \vdots \\ z_{n1} & \dots & z_{nn} \end{vmatrix}$$

Equation (F.1) now takes the form

$$(F.3) \quad dY/dx = YQ + R$$

(using matrix products, and element-by-element addition and differentiation, etc.); while the fact that each of the n rows in Z satisfies the homogeneous

equation becomes

$$dZ/dx = ZQ$$

Finally, if  $C = ||c_1 \dots c_n||$  is a set of constants, every a linear combination of the  $n$  solutions in  $Z$  with these constants as their coefficients is the one-row matrix written as  $CZ$ . The fact that it satisfies the homogeneous equation is shown by the obvious manipulation

$$d(CZ)/dx = C(dZ/dx) = (CZ)Q$$

Our object is to find a solution of the non-homogeneous equation (F.1), which vanishes when  $x = 0$  (there is only one such solution, by the uniqueness theorem). It is easily found by Lagrange's method of "variation of constants". We write  $Y = CZ$  and regard  $C$  as a row of functions of  $x$  instead of constants. Then we have

$$\begin{aligned} \frac{dY}{dx} &= \frac{dC}{dx} Z + C \frac{dZ}{dx} \\ &= \frac{dC}{dx} Z + CZQ \end{aligned}$$

In order that (F.3) be satisfied, we must have

$$\frac{dC}{dx} Z + CZQ = CZQ + R$$

So that  $(dC/dx)Z = R$ . Now since  $Z$  is a fundamental system, its determinant is never zero so that it has an inverse matrix  $Z^{-1}$ . Hence

$$\frac{dC}{dx} = RZ^{-1}$$

Therefore, we have but to take

$$C(x) = \int_0^x R(x') Z^{-1}(x') dx'$$

(x' denoting the variable of integration) and we obtain

$$(F.4) \quad Y(x) = C(x) Z(x) = \int_0^x R(x') Z^{-1}(x') Z(x) dx'$$

the required solution of (F.3).

Returning to the non-matrix form of writing, (F.4) becomes

(for  $i = 1, \dots, n$ )

$$(F.5) \quad y_i = \int_0^x [r_1(x') K_{1i}(x', x) + \dots + r_n(x') K_{ni}(x', x)] dx',$$

where  $K_{ji}(x', x)$  is the element  $[Z^{-1}(x') Z(x)]_{ji}$ ; i.e., the element in the i'th column and the j'th row in this product of square matrices. This is the form of the theorem used in Section 12, (12.8). There the six non-homogeneous terms corresponded as follows with our present  $r_i$ :

0	0	0	$\partial W'/\partial y'$	$\partial W'/\partial y'$	$\partial W'/\partial z'$
$r_1$	$r_2$	$r_3$	$r_4$	$r_5$	$r_6$

Consequently the form used had a sum of only three terms under the integral -- the perturbing terms  $\partial W'/\partial x'$ , etc; hence the equations (12.8).

APPENDIX G  
TECHNICAL MATHEMATICAL SUPPLEMENT

This appendix is addressed to the mathematical reader who requires more details concerning the existence of Fermat families, and ergodic and information theory. It may be omitted by the general reader of this report.

The Fermat family of emitted radiation. The whole treatment of waves leading up to the use of rays and the formulation of power flow in phase space has been based on the assumption of the existence of Fermat families of travelling waves, in enough cases to handle the phenomena of long range propagation. In view of the treatments in the literature [References are given in Courant & Hilbert<sup>3</sup>, Vol. II, Chap. VI, Sec. 5], the short range situation is completely understood and requires no further discussion in this place. It can be handled by rigorizing the "W.K.B. methods". On the other hand, as we have repeatedly stated, such methods require extension to cover the case when the functions become multiple-valued, as with the occurrence of caustics and multiple-path transmission. We turn, therefore, to the construction of Fermat families under the required more general conditions: emission from a point  $M_0: (x_0, y_0, z_0)$ , received at  $M: (x, y, z)$ , in a heterogeneous but isotropic medium, in which  $\rho$  and  $c$  are analytic and independent of time. We shall assume--for reasons of simplicity rather than of principle--that the physical boundaries are so far away that they can be ignored. For similar reasons we shall define and study, for any fixed locality, the finite number of wave trains from the emitter that

traverse it, corresponding with the finite number of paths. This is sufficient for the purposes of Sections 10 and 11. A global representation is not required.

As explained in the latter part of Section 6, we have to operate in space-time, and let  $t_0$  be the moment of emission from  $M_0$  of a phase which reaches  $M$  at the later moment  $t$ . It is as a function of the two sets of variables,  $(t_0, M_0)$  and  $(t, M)$  that the elementary solution  $u = u(t_0, M_0; t, P)$  of J. Hadamard<sup>3</sup> is constructed, with results that we now give in outline. The details are supplied in his "Lectures", Book II, Chapter III, under much more general conditions than we need--and with correspondingly greater limitations of range of the results.

D'Alembert's wave equation (3.1) is  $L(\psi) = 0$  where, using subscripts for partial differentiation

$$L[\psi] = \nabla \cdot (\rho \nabla \psi) - \rho n^2 \psi_{tt}, \quad n = 1/c.$$

This is an equation of hyperbolic type, and has the family of characteristic hypersurfaces (3-dimensional manifolds), exhibited as the loci, for different values of the parameter  $K$ , of  $W(x, y, z, t) = K$ , where  $W$  is a solution of the "characteristic equation" corresponding to (3.1), namely

$$2Q[W] = |\nabla W|^2 - n^2 (W_t)^2 = 0.$$

If  $(P_x, P_y, P_z, P_t)$  are any four variables, the quadratic form, whose coefficients are functions of  $(x, y, z, t)$ ,

$$Q = 1/2 (P_x^2 + P_y^2 + P_z^2 - n^2 P_t^2)$$

is called the characteristic form of the wave equation. It serves as a 4-dimensional "Hamiltonian" (cf. Section 8) to define curves, called bicharacteristics (since they are the characteristics of the characteristics of the wave equation) and containing the rays as a sub-class. The bicharacteristics are the integral curves satisfying the canonical equations

$$dx/d\sigma = \partial Q/\partial P_x = P_x, \dots, dt/d\sigma = \partial Q/\partial P_t = -n^2 P_t$$

$$dP_x/d\sigma = -\partial Q/\partial x, \dots, dP_t/d\sigma = -\partial Q/\partial t = 0,$$

where  $\sigma$  is an auxiliary parameter, the independent variable, in terms of which the eight coordinates  $(x, \dots, P_t)$  are expressed. Through each point of the 8-dimensional phase space, one and only one bicharacteristic passes; cf. Section 9.

An important fact results from the independence of coefficients of  $t$ , leading to the last equation above, which shows that  $P_t = P_t^0$ , a constant. We may exclude the case that it is zero, since then the fourth equation would give  $t = \text{constant}$ , which is an un-needed degenerate case. The sign of this constant will determine, again by the fourth equation, whether  $t$  increases or decreases with the parameter  $\sigma$ . In our study of emission from M it is convenient to have them change in the same direction, i.e.,  $dt/d\sigma > 0$ ; accordingly we take  $P_t^0 < 0$ . Note that we could then use  $t$  as the independent variable, eliminating  $\sigma$  and dropping two of our eight equations, but losing the space-time symmetry.

Our equations show that along the bicharacteristics,  $Q$  does not change its value: for evidently  $dQ/d\sigma = 0$ . There are, therefore, three classes of bicharacteristics: those for which  $Q > 0$  ("space-like"); for

which  $Q = 0$  ("ray-like"); and for which  $Q < 0$  ("time-like"). These catch-words are those used in relativity, where  $c$  is the speed of light.

By differentiating the first four equations and then eliminating the  $P$ -quantities by using the remaining four, our bicharacteristics are seen to be solutions of four differential equations of the second order,  $(x,y,z,t)$  being the dependent variables,  $\sigma$  the independent. To any set of initial values of these and their  $\sigma$ -derivatives, there corresponds a unique bicharacteristic. Now in view of the fact that on the whole length of each bicharacteristic,  $dt/d\sigma$  is not only positive but greater than a positive constant ( $-P_t^0$  times the greatest lower bound of  $n^2 = 1/c^2$ ), these curves continue indefinitely as measured by  $t$ : they never turn back. Thus each bicharacteristic emanating with increasing  $\sigma$  from the point  $(t_0, M_0)$  is cut in exactly one point by each hyperplane  $t - t_1 > t_0$ , and without being tangent to it. All this is the geometrical way of saying that  $t$  could be used instead of  $\sigma$  as independent variable.

We note in passing that our eight canonical equations for the bicharacteristics based on  $Q$  can be reduced to six based on the  $H$  of (8.4) but without space-time symmetry: we have but to eliminate  $\tau$  in favor of  $\tau$  defined by  $d\tau = c^2 dt$ , write  $P_x = -P_x/P_t$ , etc., and  $H = Q/2P$ .

The elimination of the  $P$ 's and reduction of the 8 equations to 4 of the second order in the space-time coordinates  $(x,y,z,t)$  amounts to a "projection" of phase space onto the latter:

$$(x,y,p,t,P_x,P_y,P_z,P_t) \rightarrow (x,y,z,t) = (M,t).$$

But through each  $(M,t)$  infinitely many curves which are projections of bicharacteristics may pass; and since they will have different  $P$ -values

there, their Q-value will in general be different. The characterizations of "space-like", etc., apply only to the individual curves, not to regions in space-time--contrary to the classical picture, which may only be valid locally.

The set of all rays through a given point  $(M_0, t_0)$  with  $t \geq t_0$  ("positive" rays) are the projections of those bicharacteristics through that point for which  $Q = 0$ . They form what is called the characteristic conoid, of vertex at  $(M_0, t_0)$ , and correspond with the paths of signals emitted from  $M_0$  at the epoch  $t_0$ . They form a 2-parameter family of curves in 4-space-time, and a 1-parameter in 3-space-time. Therefore the conoid is a 3-dimensional manifold in the former and a 2-dimensional (surface) in the latter. In all cases it is a characteristic surface of the wave equation; i.e., it is given (locally at least) as the locus of  $W(M, t) = K$ , where, on replacing  $\partial W / \partial x$ , etc., by  $P_x$ , etc.,  $Q(P_x, \dots) = 0$ .

The picture of the conoid is simplest when  $c$  is constant (or nearly so). Then in the case of  $(x, y, t)$ , the conoid is a cone of revolution about an axis parallel to the  $t$ -axis, vertexed at  $(x_0, y_0, t_0)$ , and of semi-conical angle of tangent equal to  $c$  (or when  $c$  varies slightly, a slight deformation of this figure). The projection onto the  $xy$ -space has no vacant spaces and no multiple paths: this is the classical picture. As we have just seen, it may lose its validity if  $c$  is more variable and the extent of space considered greater.

A regular point of the conoid is one at which it can be represented locally as  $W = K$ , where not all four first derivatives of  $W$  vanish. At such a point it is smooth; and it can have no tangent line parallel to the  $t$ -axis, since this would imply that  $\partial W / \partial t = 0$ , and hence, by the



characteristic equation satisfied by  $W$ , that the sum of squares of the other derivatives vanishes, so that the latter are individually zero. Irregular points (where such a representation is not possible) are either points where two smooth pieces of the conoid intersect (each of which can be represented as above but with different functions  $W$ ) or else edges of regression, i.e., envelopes on the conoid of its generating rays. In either case the dimensionality of the locus of singularities is one lower than the space of the medium--two lower than the space-time. These statements follow from the general theory; they may have exceptions, but in the analytic case assumed, these will be still lower in dimensionality (e.g., when the singularity is a conical point).

We return to the projection of the characteristic conoid vertexed at  $(M_0, t_0)$  and its generating rays, upon the space of the medium. A point  $M$  of the latter may be said to be of "multiplicity  $m$ " ( $= 0, 1, 2, \dots$ ) if just  $m$  rays connect it with  $M_0$ ;  $m = 0$  represents a point in a vacant region. Let  $M_1$  be a boundary point of regions of different multiplicity: clearly a vertical line through  $M_1$  touches the conoid, and since it cannot be tangent to it at a regular point (or at the intersection of two smooth pieces), it must touch it at a point of an edge of regression. Thus the edges of regression of the conoid project into the caustics in the space of the medium--these being the boundaries of regions of different multiplicities of the latter.

To picture an edge of regression, we have but to take any smooth twisted curve  $C$  in  $(x, y, t)$  whose direction always makes an angle with the vertical, the tangent of which angle equals the value of  $c$  there (it will be a solution of the Monge equation  $dx^2 + dy^2 - c^2 dt^2 = 0$ ).

A small characteristic conoid vertexed at a point of  $C$  will be tangent to it; and as its vertex moves along  $C$  its two parts will envelop a surface having  $C$  as edge of regression, and being a characteristic surface of the wave equation: the conoid itself, if  $C$  is chosen properly. See Figure G-I.

Let  $M$  be any point inside a region of multiplicity  $m > 1$ : just  $m$  paths join it to  $M_0$ . Let  $t', t'', \dots, t^{(m)}$  be the epochs at which a signal starting from  $M_0$  at the epoch  $t_0$  reaches  $M$ : they are in general all different; and since  $M$  is not taken on a caustic, can cease to be so only if  $M$  is the projection of a point of intersection of two or more smooth pieces of the conoid. As we have seen, such exceptions are confined to loci of lower dimensionality than the space of the medium--as remarked late in Section 6. Moreover, they correspond to  $m$  different analytical elements. Indeed, if  $t' = t''$  at  $M = M_1$ , then the functions  $t'(M)$  and  $t''(M)$ , with  $M$  in a neighborhood of  $M_1$ , can be expanded in convergent powers of the differences of the coordinates of  $M$  and  $M_1$ : as power series they are quite different; they merely happen to take on the same value at  $M_1$ .

Quite different is the behavior of the multiple-valued function  $t = t(M)$  near a point  $M_1$  on a caustic: the latter is a branch locus of this function, one branch at least being non-analytic there (compare the behavior of the function  $t$  defined by  $t^2 = x^3$  in the 1-dimensional medium of  $x$ ). Since the eikonal function  $S$  of Sections 1.4, etc. is the above  $t(M)$ , the mechanism of its multiple-valuedness is exhibited in this space-time figure.

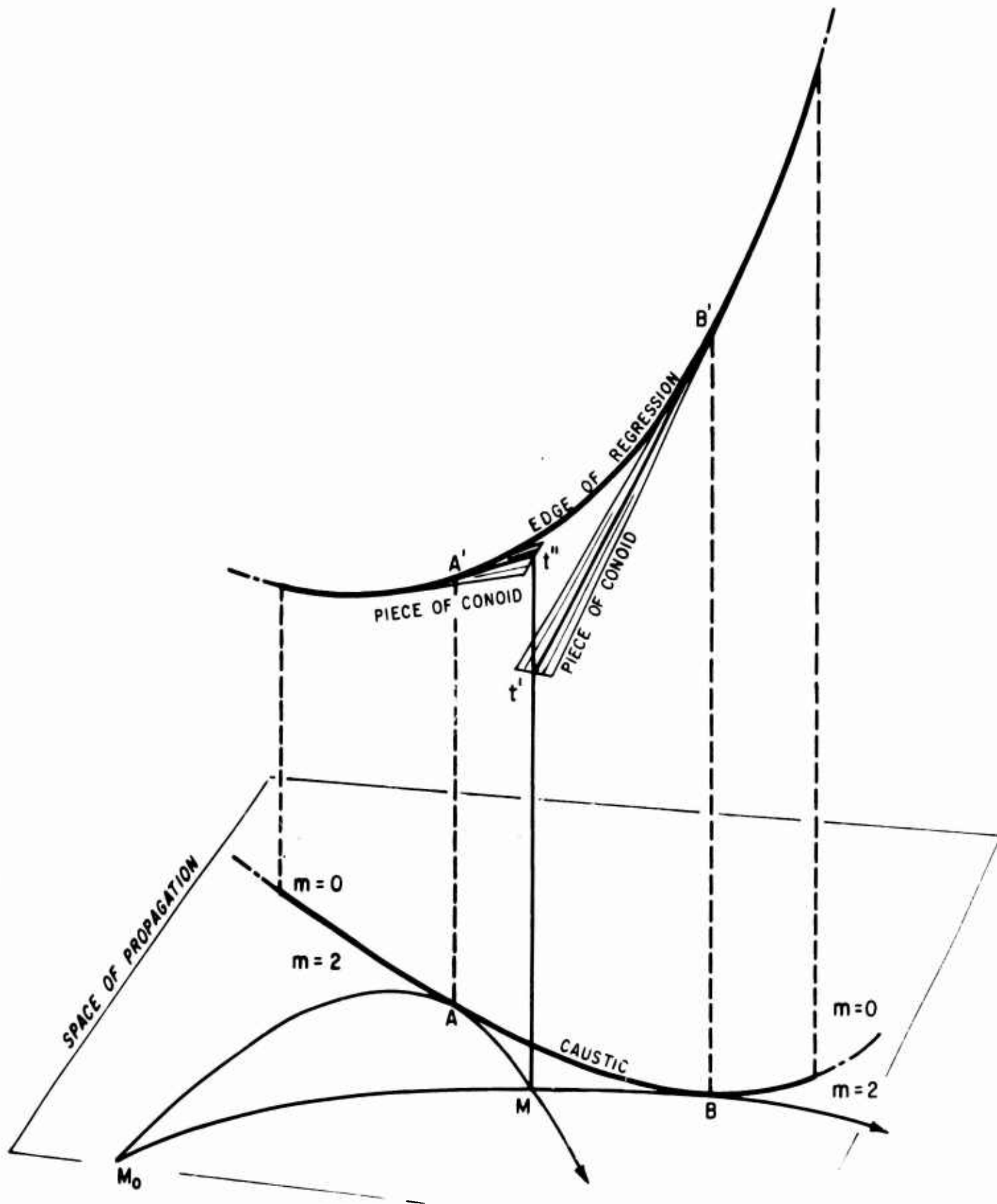


FIGURE G-1. MULTIPLE PATHS AND VALUES OF  $t$ .

Hadamard's elementary solution  $u$  is based on a function  $\Gamma$  of eight variables, the space-time coordinates of "initial point"  $(t_0, M_0)$  and of the "final point"  $(t, M)$  on a bicharacteristic:  $\Gamma$  is the square of the geodesic distance  $G$  between them in space-time, based on the metric defined by the element of arc

$$dG^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2$$

associated with the wave equation. The elementary formalism of the calculus of variations shows that the Euler equations for an extremal of  $dG$  coincide with the second order system in  $(x, y, z, t)$  derived as noted above from the canonical equations. Hence the geodesics in space-time are the bicharacteristics, and we easily calculate that the element  $dG$  along them is given by  $\sqrt{2Q} d\sigma = \sqrt{2Q_0} d\sigma$ , since  $Q$  is constant along each line. Hence, the geodesic distance is  $\sqrt{2Q_0} \sigma$ ,  $\sigma$  being the value of the parameter at the final point, it having initiated at zero. Hence,  $\Gamma = G^2 = 2Q_0 \sigma^2$ , so that  $\Gamma > 0$  for space-like bicharacteristics,  $\Gamma < 0$  for time-like ones, and  $\Gamma = 0$  for rays: this is the equation of the characteristic conoid. A standard theorem in the calculus of variations is that  $|\nabla G|^2 - n^2 G_t^2 = 1$ ; it follows at once that  $\Gamma$  satisfies the first order partial differential equation  $|\nabla \Gamma|^2 - n^2 \Gamma_t^2 = 4\Gamma$ . A crucial advantage in using  $\Gamma$  instead of  $G$  is that it is an analytic function of its eight variables, whereas  $G$  is not. [References are in Hadamard, l.c.; this  $\Gamma$  has no relation to our earlier usage of the same letter; and our present use of  $G$  is not Hadamard's].

The elementary solution  $u$  has a drastically different form according to whether the space-time is of odd or even dimensionality.

In the latter case there is a logarithmic term, absent in the odd case. In radiation along a line (2-dimensional space-time)  $u = V \log \Gamma$ , where  $V$  is a special analytic solution of the wave equation: this is Riemann's function. In radiation in a plane  $u = V/\sqrt{\Gamma}$ , while in the case of spacial propagation, of primary concern to this study\*, we have

$$u = U/\Gamma - V \log \Gamma$$

where  $U$  and  $V$  are analytic, and  $V$  satisfies the wave equation. Note that when  $c$  is constant,

$$\Gamma = (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 - c^2(t-t_0)^2, \quad V = 0, \quad U = \text{constant},$$

and we have the familiar expression  $u = 1/\Gamma$ .

There is a special feature of importance in the present case of time-independent coefficients of our wave equation: any solution will remain a solution if any constant is added simultaneously to both  $t$  and  $t_0$ : this means that these variables enter  $\Gamma$  only through their difference  $t - t_0$ . Therefore  $\Gamma = \Gamma(t-t_0, M_0, M)$ . Moreover, on following through the steps of Hadamard's construction, it can be shown that the same may be assumed for  $u$ ,  $V$ , and  $U$ . Furthermore, since for fixed values of  $(x_0, y_0, z_0, t_0)$ ,  $u$  satisfies the wave equation in  $(x, y, z, t)$ --provided  $\Gamma \neq 0$ --we can multiply  $u$  by a regular function of the subscripted variables and integrate over any path of the latter, provided points of  $\Gamma = 0$  are avoided. We shall make use of these facts to construct a

\*While we are very much concerned with the special case of "radiation in a vertical plane" we really mean "in space with azimuthal symmetry"--an absolutely different situation from the case of a medium enclosed between perfectly rigid parallel and vertical walls, or from an infinite line radiator, etc.

set of local Fermat families, having a total effect of asymptotic families, according to the broadened definitions at the close of Section 6.

First it is essential to note that in all cases of interest to the present report the geodesic length  $G$  and its square  $\Gamma$  are multiple-valued--even in the simplest case of laminarity and a single duct (minimum of  $c$ ). This is an essential part of the multi-path situation. Its analogue in geodesics on curved surfaces is illustrated in cases in which two points on the surface can be joined by many different geodesics, having different lengths.

To understand the situation in our space-time picture, we may observe that the first order partial equation satisfied by  $G$  mentioned above  $2Q = 1$  when solved for  $G$  by the method of characteristics [Cf. Courant-Hilbert<sup>3</sup>, Vol. II, Chap. II, Sec. 7 et sec.] gives the five quantities  $(x,y,z,t,G)$  as single-valued regular functions of four variables: the parameter ( $\sigma$  or  $s$ , etc.) along each characteristic, and three others, which in our case (the initial values of the coordinates being held constant and only the momenta  $P$  varied) would be three variables specifying the position of the momentum point on the manifold  $2Q = 1$ . In other words, it gives  $(x,\dots,G)$  as single-valued functions of a point  $(\sigma,P)$  on the Cartesian product of the real axis and this quadratic manifold. Under the traditional limitation to local assumptions, there is no difficulty in inverting the  $(\sigma,P) \rightarrow (x,y,z,t)$  map, thus expressing  $(\sigma,P)$  as single-valued functions of  $(x,y,z,t)$ , which are then substituted in place of  $(\sigma,P)$  in the fifth expression  $G = G(\sigma,P)$ , which thus yield the single valued result  $G = G(x,y,z,t)$ . Of course, it is impossible in the large to invert the  $(\sigma,P) \rightarrow (x,y,z,t)$  map:  $\sigma$  and  $P$  are multiple-

valued functions of the coordinates, a multiple-valuedness which transmits itself to  $G$  as a function of them. Similarly,  $\Gamma = G^2$  is multiple-valued.

The fact--derived directly from the general theory--that with our assumptions of analyticity  $(x,y,z,t,\Gamma)$  are analytic and single-valued in terms of  $(\sigma,P)$  shows just what kind of branching loci enter in the multiple-valuedness: being given by analytic equations (the vanishing of certain Jacobians formed from the above analytic functions) they are regular analytic manifolds of dimensionality one less than that of the space-time (3, in the above case). Therefore we can consider the different single-valued analytic branches of  $\Gamma$ --as we shall now do.

An examination of Hadamard's construction shows that his elementary solution is valid locally (near the point  $(x_0, y_0, z_0, t_0)$ ) with  $U$  and  $V$  single-valued and analytic there. These and all their properties except single-valuedness are extended by analytic continuation to the regions of present concern.

We return to the construction of the Fermat families, confining attention now to regions of our space-time away from the branch loci, and to individual single-valued branches. In this sense, we continue to use the expressions  $\Gamma$ ,  $u$ ,  $U$  and  $V$ . We shall also operate in regions of space-time away from edges of regression of the characteristic conoids, projecting into the space of the medium away from caustics. Our purpose is to derive, from the instantaneous pulse emitted from  $(M_0, t_0)$  represented by  $u$ , a continuous harmonically modulated train of waves.

Let  $M$  be any fixed point in such a region of the medium, and  $t'$ ,  $t''$ , etc., the points of time at which the line through  $M$  and parallel to the  $t$ -axis cuts the conoid vertexed at  $(M_0, t_0 = 0)$ . Cf. Figure G-I.

Since  $\Gamma$  depends on  $t, t_0$  only through the difference  $t - t_0$ , the conoid can be raised or lowered by the amount  $t_0$ , and we still have  $\Gamma(t-t_0, M)=0$  for  $t$  fixed but  $t_0$  taking on the values  $t'_0 = t-t'$ ,  $t''_0 = t-t''$ , etc.

We have seen that these solutions  $t', t''$ , etc. of the equation  $\Gamma(M, t) = 0$  are analytic functions of the coordinates of  $M$ ; we shall denote them by  $S = S(x, y, z)$ , the multiple-valued eikonal function; indicating, when necessary, its separate branches by accents. Since near  $(t', M)$   $\Gamma(M, t)$  is analytic and vanishes when  $t-S = 0$ ; and since  $\partial\Gamma/\partial t \neq 0$  there (no vertical tangent line to the conoid), Weierstrass' factorization theorem tells us that  $\Gamma(M, t) = (t-S)A$ , where  $A$  is an analytic function of  $(x, y, z, t)$  in a neighborhood of  $(M, t)$ . It is convenient (although not strictly necessary) to assume further that  $M$  has been chosen not to be on any 2-dimensional locus of coincidence  $t' = t''$ , etc. Then  $A$  is not zero in the neighborhood of  $(M, t)$ . On the basis of these assumptions and results, we proceed as follows.

We get the general values of our functions by replacing  $t$ , when it occurs alone in the above, by  $t-t_0$ ; we have

$$\Gamma = \Gamma(t_0, M_0; t, M) = (t-t_0-S) A(t-t_0, M_0, M)$$

Naturally we are confining all the variables to the above neighborhood.

We now multiply  $u$  by  $(1/2\pi i) \exp(-i\omega t_0)$  and, regarding  $t_0$  as a complex variable, integrate the product in the complex  $t_0$ -plane about circles centered at the points  $t'_0, t''_0$ , etc., and with sufficiently small radii to exclude any possible irregular points of the functions. The result is a sum of as many integrals as there are  $t_0$ -values obtained by the intersection process described. Changing variables of integration, by setting  $Z = t-t_0-S$ ,  $dZ = dt_0$ , whence  $\exp(-i\omega t_0) = \exp[i\omega(S-t)] \cdot \exp(i\omega Z)$ , our integral becomes, on taking outside the integral the factor independent of  $Z$

$$\psi' = \frac{1}{2\pi i} \exp[i\omega(S-t)] \int u \exp(i\omega Z) dZ$$



Locally, this has the form of a classical travelling wave of frequency  $\omega$ , since the integral will depend on  $\omega$  and the space coordinates only, and not the time. This is because the integrand  $u = u(t-t_0, M_0, M) = u(Z + S(M_0, M), M_0, M)$ ; and  $Z$  is integrated out. But we must examine the behavior of the terms with increasing  $\omega$ .

The term  $U/\Gamma$  in  $u$  is found at once by applying Cauchy's formula, since the denominator is

$$\Gamma = (t-t_0-S) A(t-t_0, M_0, M) = Z A(Z + S, M_0, M)$$

while  $U = U(t-t_0, M_0, M) = U(Z + S, M_0, M)$ . Thus the integral, which we denote by  $R'$ , is

$$R' = \frac{1}{2\pi i} \int \frac{U}{A} \exp(i \omega Z) \frac{dz}{Z}.$$

This, by Cauchy's integral formula is the value, for  $Z = 0$ , of the coefficient of  $dZ/Z$ , namely

$$R' = \frac{U(S(M_0, M), M_0, M)}{A(S(M_0, M), M_0, M)}$$

An essential fact shown by this is that  $R'$  is independent of  $\omega$  and is a function merely of the space coordinates of  $M_0$  and  $M$ . The same was clearly true of  $S$ . Furthermore, all these functions are real for real values of the variables (as is easily seen by their definitions). Consequently this contribution to  $\psi'$  is  $R' \exp[i \omega (S-t)]$  and  $R'$  and  $S$  are real valued functions of position, independent of time and frequency. In fact, if this computation is applied to the case of homogeneous media and the corresponding simple form of  $\Gamma$  given above, we immediately get the same point radiator formula given above in our fourth example of Appendix D.

A corresponding derivation of our third example of the point radiator in the plane could be produced, starting from  $u = 1/\sqrt{\Gamma}$ ,  $\Gamma = (x-x_0)^2 + (y-y_0)^2 - c^2(t-t_0)^2$ ; only we would have to integrate about an appropriate closed path in the two-sheeted Riemann surface for the complex square root, and then apply standard contour integral expressions for the Hankel functions.

Returning to the general case, we must now treat the logarithmic term in  $\psi'$ . We have, after the above change of variable of integration,

$$-V \log \Gamma = -V \log A + V \log Z$$

Since the first term is analytic in the neighborhood of our path of integration, its integral vanishes, by Cauchy's theorem. We are left then with the integral of the product of the analytic function, which we write for simplicity as  $f(Z) = V \exp(i \omega Z)$ , with the factor  $\log Z$ , about a small circle of radius we may denote by  $r$ . Let  $r'$  be the radius of a smaller concentric circle in the  $Z$ -plane, and consider the region  $\Sigma$  bounded by the two circles and the radial line  $(r', r)$  on the axis of reals; see Figure G-II. In this  $\Sigma$  we take the determination of the logarithm with an imaginary part approaching zero as  $Z$  approaches the radial line from above; and therefore approaching  $2\pi i$  as this line is approached from below. Since throughout  $\Sigma$  the function  $f(Z) \log Z$  so defined is analytic, Cauchy's theorem tells us that its integral about the whole boundary, followed in the directions shown by the arrows, vanishes. Hence, the integral around the larger circle--which we want to find--is that about the smaller circle, plus the difference between the integrals along the radial line of the two determinations of

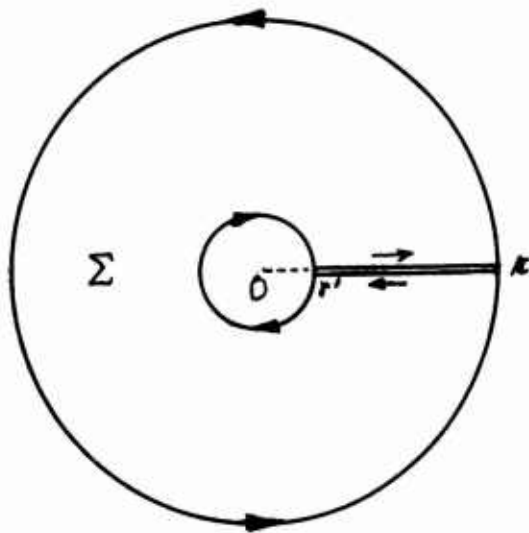


Figure G-II. Integration in the Complex Z-plane

$f(z) \log z$ ; in other words, of  $2\pi i f(Z)$ ; and since  $Z = X$  on this line, this is  $2\pi i \int_{r'}^r f(X) dX$ . Finally, on the small circle, an elementary inequality shows that it is less in absolute value than a constant times  $r' \log r'$ , and hence approaches zero as  $r' \rightarrow 0$ . Therefore the contribution of our logarithmic term is  $\int_0^r f(X) dX = \int_0^r V(X) \exp(i\omega X) dX$ .

Now  $V(X)$  being analytic in its variables (all but  $X$  not being written explicitly), we may integrate by parts, obtaining

$$\begin{aligned} \int_0^r V(X) \exp(i\omega X) dX &= \frac{1}{i\omega} \int_0^r V(X) d \exp(i\omega X) \\ &= \frac{1}{i\omega} \left\{ [V(X) \exp(i\omega X)]_0^r - \int_0^r V'(X) \exp(i\omega X) dX \right\} \end{aligned}$$

and the quantity in  $\{ \}$  is easily shown to be bounded with increasing  $\omega$ . Therefore we have proved that the contribution of the logarithmic term in  $u$  approaches zero (and locally uniformly) as  $\omega \rightarrow 0$ . By the analyticity, the same will be true of its derivatives.

This is enough to show--by obvious computations of  $\vec{F}$  and  $E$ --that the one wave  $\psi'$  corresponding to the path which is the projection into space of the bicharacteristic connecting  $M_0$  with the  $t'$  intersection on the conoid (see Figure G-I) is in fact a Fermat family. But with many intersections we have multiple paths, and our wave function  $\psi = \psi' + \psi'' + \dots$ ; in view of Theorem II of Section 6, this would not be a Fermat family, because of the interference terms in the products.

It is at this point that the extension of the definition, noted at the close of Section 6, has its application: the replacement of "strong" limits by "weak" ones--i.e., in averages. In the products of two first order partial derivatives occurring in the calculation of  $\vec{F}$  and  $E$ , if at least one factor comes from a logarithmic part of the elementary solution, the product will approach zero (strongly!) as  $\omega \rightarrow \infty$ . The cross products of concern come from the non-logarithmic parts, e.g., from  $R' \exp[i\omega(S'-t)]$  and  $R'' \exp[i\omega(S''-t)]$ . Now the interference terms contain as factors sines and cosines of  $\omega(S'-S'')$ , multiplied by analytic coefficients that are independent of  $\omega$  since they come from  $R$  and  $S$  functions and their derivatives: they are of the form  $f(x,y,z) \cos[\omega g(x,y,z)]$ , or with a sine replacing the cosine.

That the integral of such an expression over any small region (not containing irregular points of these functions) approaches zero as  $\omega \rightarrow \infty$  is shown as follows: If  $g$  is a constant, the two waves differ only by this fixed phase factor and they can be written as a single wave. If  $g$  is not a constant, being analytic it must have at least one non-vanishing derivative except possibly on lower dimensional loci--which can be avoided, since they produce but an arbitrarily

small contribution to the integral. Suppose that, e.g.,  $g_x$  is not identically zero, and take our region of integration such that its absolute value is greater than some positive constant. On writing

$$f \cos \omega g = \frac{1}{\omega} \left[ \frac{\partial}{\partial x} \left( \frac{f}{g_x} \sin \omega g \right) - \sin \omega g \frac{\partial}{\partial x} \left( \frac{f}{g_x} \right) \right]$$

and applying the divergence theorem to its integral over our small region, we show that the bracket leads to bounded volume and surface integrals; and so the factor  $1/\omega$  makes the product approach zero as  $\omega \rightarrow \infty$ : this shows the "weak" convergence to zero of the interference terms.

To sum up what we have accomplished: we have shown that in multipath transmission from the emitter to the neighborhood of the receiver, as many trains of waves (functions satisfying the wave equation, etc.) as there are paths exist, each one of which satisfies, in this neighborhood, the conditions required by a Fermat family; and that their total effect is, in terms of weak convergence, an asymptotic family, the energy and power flux being (vectorially) additive. This is what was required in Sections 10 and 11--in which the discussion was put into an intuitively geometrical form.

It is hardly necessary to point out to the mathematical reader the major incompletenesses of this discussion: not only in carrying out many details, but in showing that a global, single-valued, wave  $\psi$  can be found by uniting our separate paths of integration into one in the Riemann multiple-space (general R. surface) upon which  $\Gamma$ ,  $u$ , and the related functions are single-valued. Exact hypotheses would be needed, etc.; but this would go far beyond the scope of this report: let the above outline call attention to the basic incompleteness of the classical treatments and serve as a stimulus to further work!

The Ergodic Theory. This theory has been cited in Section 14, as a going beyond the intuitive picture of mixing of regions in the surface of section, appealed to in deriving (14.6). It has been cited again in Section 16, in the more rigorous second method, in which the secular non-laminar case was transformed canonically into a situation similar to the laminar one, in which the ergodic theory could apply. As already noted, this application involves the replacement of the actual situation as it occurs at long but finite ranges by an idealized limit as the range increases indefinitely. The justification is--as usual in such replacements of large finite situations by limiting ones in physics--in the assumption that a satisfactory approximation is thus obtained.

The ergodic theory deals with a one-parameter group of measure-preserving transformations of a region of finite measure onto itself. The parameter was  $\tau$  (or  $\tau$ ) in Section 14, and  $\sigma$  in 16. The region transformed was the part of the surface of section inside a Snell curve corresponding to the duct under consideration. The measure was the area of any chosen piece of this surface.

The theory has a long history, starting with Boltzmann and Maxwell's "Ergodic Hypothesis" in statistical mechanics, its modification by P. and T. Ehrenfest's "Quasi-Ergodic Hypothesis", and the ideas of J. W. Gibbs, formulated by H. Poincaré as the "Ink Spot Problem".<sup>6</sup> These authors simply hypothesized, in increasingly precise forms, the results of applying intuition to the picture of mixing--as we have

described it in these pages. The first set of proofs came in the early 1930's: v. Neumann's use of the present author's treatment of Hamiltonian systems by means of Hilbert space<sup>6</sup> to prove the "Mean Ergodic Theorem",<sup>6</sup> G. D. Birkhoff's subsequent proof of the "Ergodic Theorem",<sup>6</sup> followed by work of these authors and E. Hopf on the relations of this theory to the "Ink Spot Problem". The history of the subject up to 1932, as well as a statement of the theorems (without proofs), is contained in G. D. Birkhoff and the author's joint note "Recent Contributions to the Ergodic Theory".<sup>6</sup> During the ensuing forty-three years the activity in the pure mathematics suggested by these investigations has become enormous and is continuing in great volume. Since, however, it has for the most part not been oriented toward physical applications, and is not needed for the present ones, it is left out of account here.

The formulation of the very restricted form of this theory needed in Sections 14 and 16 is discussed in simple terms but with complete mathematical rigor by P. R. Halmos<sup>6</sup> in his chapter entitled Mixing.

Information Theory. The application of the modern theory of information to the discussion of random effects in Section 15--as well as its implied possibility in the more general situation of Section 16--reposes on definitions and theorems based on the convexity of certain functions and functionals. The material in the literature is now very extensive; we give but one reference, to Kullback.<sup>6</sup>

## APPENDIX H

### PROPAGATION OVER UNDERWATER OBSTRUCTIONS AND THE "COUPLING FACTOR" FALLACY

Propagation over underwater obstructions. A perennial problem in long range propagation of sound in the ocean is the quantitative appraisal of the obstructive effects of geological formations such as sea mounts and ocean ridges. When the ranges are long, any reflective effects of such obstructions disappear, and they can be regarded as pure absorbers: their effects are those of a very complicated and incompletely charted absorbing surface that forms the lower boundary of the medium propagation.

The object of this appendix is to apply the ray methods, as set forth in Sections 13 and 14, to the case of what may be called generically "sea mounts": completely absorbing geological projections from the ocean bottom that are limited in their dimension along the path of propagation to less than an average convergence zone. Their effects on propagation loss are sufficiently illustrated by the situation in which they protrude into part of a deep sound channel in the laminar case ( $c = 1/n$ , dependent on  $z$  only). As we know, the differential equations of the rays can then be solved by a single quadrature, and each ray can be characterized by the vertical plane in which it lies and by its Snell constant  $k$ , the identifying parameter that can be thought of as the reciprocal sound speed at its highest (or lowest) point; c.f. Appendix B and Section 13.



FIGURE H-1. EFFECT OF SEA MOUNT  
IN THE MEDIUM AND IN THE SURFACE OF SECTION

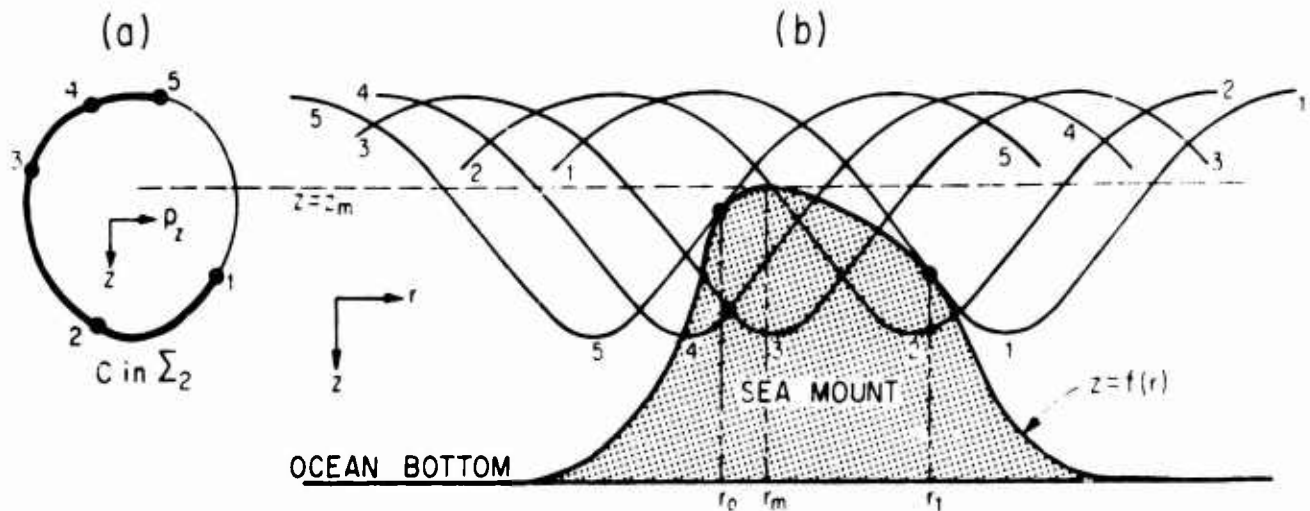


Figure H-1 shows in (b) a vertical plane section of a sea mount, together with a set of typical acoustic rays in this plane, all having the same Snell constant. The plane section may be one of a set of coaxial planes drawn through a point emitter, in the simple case of cylindrical spreading. Or it may be one of a set of parallel planes in the "plane wave" idealization. In any case, our first problem is to find how much power the sea mount absorbs from this class of rays; that is, from all the rays having this common Snell constant. The oval shown in (a) will be explained after the scheme of evaluating the sea mount's effects has been stated.

The method is simplicity itself: If all these rays cut the sea mount, none of their power gets through. If some of them do not cut it, the following construction shows the fraction of power it intercepts: All the

rays for the given Snell constant being congruent, they are obtained by moving one of them (as a rigid wire) horizontally. Let the wire be moved to the left to the position (1) at which it first touches the sea mount, through the intersecting positions such as (2), (3), (4), to the position (5) where it touches the mount for the last time. If the distance of displacement from (1) to (5) is  $R_m$  and the ray's wave-length is  $R$ , then the fraction of power intercepted by the mount is their ratio,  $R_m/R$ . Thus after passing this obstruction, the power becomes multiplied by the factor less than unity:  $1 - R_m/R$ .

To find the total effect of the sea mount, we subdivide the class of rays into sets belonging to narrow Snell constant intervals, and then apply the above construction to a representative ray in each set. The appropriate reduction factor is applied to the power in each interval, and the results are summed over all intervals.

If the obstruction has many humps, a similar construction is applied; but the intersecting positions may have several pieces: one still takes the ratio of their total length to the wave-length as the absorption factor.

In the case of widely separated obstructions, it would be unrealistic to assume strict laminarity throughout their whole range--or indeed that the distances are accurately known. Then one treats them as uncorrelated absorbers, applying each power decrement factor so as to obtain a product. In the case of very long ranges, traversing many obstructions, these repeated multiplications lead to an exponential approximation in which the obstructions are averaged into a single parameter: decibelwise, an effect proportional to range is obtained imitating a volumetric attenuation, but possibly Snell-constant-dependent.

In treating the problem, as we have, by simple geometry, we have incurred a logical debt: we must prove that what we have done is in fact a correct deduction from the wave theory of propagation. After all, other equally simple and plausible geometrical constructions have been given, which have turned out to be quite wrong--i.e., inconsistent with the wave theory.

The validity of the above construction follows from Sections 13 and 14 and is easily illustrated by reference to Figures 14-I and III of the latter. We are now dealing with the case in which the depth function  $z = b(a)$  is no longer constant but decreases to a minimum at the height of the sea mount and then returns to its constant value  $b$  used throughout Section 14.

On comparing Figure H-I(a) with Figure 14-III, which gives the surface of section in the case under consideration, we note that I(a) exhibits as an oval a locus of points in the latter for a fixed value of the Snell constant  $k$ . That part of the oval (1, 2, 3, 4, 5) which corresponds to the rays which cut the sea mount, such as 1, 2, 3, 4, 5, of I(b), is indicated by deeper shading. Now we have seen that, at the ranges considered, the power flux density is a function  $F(k)$ , and is therefore the same at each point of the oval. Hence the total power that would (in the unobstructed case) cross the narrow ring between the ovals for the given  $k$  and for  $k + \Delta k$  is  $F(k)$  times the area of the ring (to quantities of higher order in  $\Delta k$ ), which area (to the same order of approximation) is the derivative of the total area within the oval--the  $A(k)$  of Section 14, times  $\Delta k$  -- given by (14.16), which shows that it is

$R(k)\Delta k$ , where  $R(k) = R$ , the ray period shown in Figure G-I. By similar formulas based on those of Section 14, the area of any part of the ring contiguous to any arc of the oval is the corresponding horizontal distance (e.g.,  $R_m$ ) through which the ray moves as its point of intersection with the oval traces this arc. Cf. also Appendix B. Hence in the case in which the sea mount obstructs rays through such an arc, the fraction of power lost is the ratio of arcs (the factors  $\Delta k F(f)$  cancelling) and is therefore  $R_m/R$ , as stated. Of course this ratio must be considered for each value of  $k$  and the results combined by integration.

The "coupling factor" fallacy. When a signal source at A injects sound into the water, which is received at a distant receptor B, the amount of the received power--or rather, the proportion of that received to that emitted  $P_B/P_A$ --depends on the whole nature and condition of the physical system: emitter-medium-receptor. The effect of the medium--the "transmission loss" in the strict sense, has been the main subject of the present Report. As explained in Sections 11 and 13, this is defined as the ratio  $I_1/I_0$  of the intensities (power flux per unit area in the medium without regard to exact ray directions) at the respective points.

Now we have shown that under a given glass of observationally similar conditions, this ratio depends not only on the range  $r$  between A and B but on their depths  $z_0$  and  $z_1$  as well. This dependence is essential, even after much fine-structure and effects of random have been averaged out, as in the "second approximation" mentioned in the Introduction and developed in Sections 14, 15, and 16 (dependence on profiles at A and B but not on intermediate ones). The nature of the dependence can be

exhibited by the shape of the ovals on the surface of section at these two points, with the consequent drastic variation of the injection and reception horizontal strips. Cf. Figures 14-V and 16-I. Let us represent this transmission loss as  $I_1/I_0 = L(r, z_0, z_1)$ . Such a depth dependence is in fact an observational truism.

Returning to the over-all ratio  $P_B/P_A$ , we can write it as the product of three factors

$$P_B/P_A = (I_0/P_A) (I_1/I_0) (P_B/I_1)$$

The factor  $I_0/P_A$  represents the effectiveness with which the emitter A creates, by its strength  $P_A$  as a source, an acoustic intensity  $I_0$  at a reference point in its neighborhood. Under normal degrees of physical variation,  $I_0/P_A$  is independent of the actual position of A, and can be regarded as determined by its physical nature as an emitter: it may validly be called the emitter's coupling factor  $C_A = I_0/P_A$ . Similarly for  $C_B = P_B/I_1$  (or if we insist on symmetry,  $I_1/P_B$ ; but this will have no effect on the argument). Consequently we have

$$P_B/P_A = C_A L(R, z_0, z_1) C_B.$$

Now the transmission loss is frequently regarded as a function of  $r$ :  $L(r)$ , independent of  $z_0$  and  $z_1$ , whereas the observationally evident dependence of  $P_B/P_A$  on the depths  $z_0$  and  $z_1$  is ascribed to the effect of depth on the degree of coupling, and thus our  $C_A$  and  $C_B$ , defined physically as we have done above, are replaced by the functions  $C_A(z_0)$  and  $C_B(z_1)$ ;

and hence we obtain the identity in  $r, z_0, z_1$ :

$$P_B/P_B = C_A L(r, z_0, z_1) C_B = C_A(z_0) L(r) C_B(z_1).$$

This is a mathematical possibility if and only if  $L(r, z_0, z_1)$  is a product of three functions: of  $r$  alone,  $z_0$  alone, and  $z_1$  alone (to show that this follows from the above equation, divide it through by the product  $C_A C_B$ ). Since all the results of the present Report show that this is incorrect--even as a crude approximation--the above use of depth-dependent coupling factors with depth-independent transmission loss is an error. If, in spite of this fact, we take  $10 \log_{10}$  of the formulas, the decibel measure of the loss of strength from signal to receptor becomes a sum of (negative) terms: the decibel measures of the two couplings, and the classical expression of the transmission loss as range dependent only. Such types of combinations of physical effects by adding their measures in decibels, while very convenient and justified in many engineering studies, is not possible in the present case.